DEVELOPMENT AND VALIDATION OF ALGORITHMS FOR MRI SIGNAL COMPONENT ESTIMATION

by

M. Dylan Tisdall

Hons. BMath, University of Waterloo, 2002

A THESIS SUBMITTED IN PARTIAL FULFILLMENT
OF THE REQUIREMENTS FOR THE DEGREE OF
DOCTOR OF PHILOSOPHY
in the School
of
Computing Science

SIMON FRASER UNIVERSITY
Fall 2007

APPROVAL

Name: M. Dylan Tisdall

Degree: Doctor of Philosophy

Title of thesis: Development and Validation of Algorithms for MRI Signal Component Estimation

Examining Committee: Dr. Torsten Möller
Chair

Dr. Stella Atkins, Senior Supervisor

Dr. Ghassan Hamarneh, Supervisor

Dr. Mike Hayden, Supervisor

Dr. Bob Russell, Supervisor

Dr. Richard Lockhart,
Professor of Statistics amd Actuarial Science
SFU Examiner

Dr. Anne Martel, External Examiner,
Assistant Professor of Medical Biophysics
University of Toronto

Date Approved: December 7, 2007

SFU SIMON FRASER UNIVERSITY
THINKING OF THE WORLD

STATEMENT OF ETHICS APPROVAL

The author, whose name appears on the title page of this work, has obtained, for the research described in this work, either:

(a) Human research ethics approval from the Simon Fraser University Office of Research Ethics,

or

(b) Advance approval of the animal care protocol from the University Animal Care Committee of Simon Fraser University;

or has conducted the research

(c) as a co-investigator, in a research project approved in advance,

or

(d) as a member of a course approved in advance for minimal risk human research, by the Office of Research Ethics.

A copy of the approval letter has been filed at the Theses Office of the University Library at the time of submission of this thesis or project.

The original application for approval and letter of approval are filed with the relevant offices. Inquiries may be directed to those authorities.

Bennett Library
Simon Fraser University
Burnaby, BC, Canada

Abstract

The MRI analysis pipeline consists of a data-acquisition stage defined by a protocol, an estimation stage defined by a function, and an analysis stage – normally performed by a radiologist. MRI data is acquired as a 3D or 4D grid of complex-valued measurements. In some protocols more than one set of measurements are fused into a vector of complex values. However, radiologists normally desire a real-valued 3D or 4D dataset representing a feature of interest. To convert from the measurements to the real-valued feature an estimator must be applied.

This thesis studies the development and evaluation of estimators. We approach the problem not as one of general image processing, but as one specific to MRI and based in the physics of the measurement process. The estimators proposed are based on the physics of MRI and protocols used clinically. We also show how estimators can be evaluated by testing suitability for radiological tasks.

We present statistical models for protocols and features of interest that arise in MRI. Since the models contain nuisance parameters many estimators are available from the statistical theory. Additionally, we consider how adding a constraint of regularity in the phase coordinate of the complex data affects the estimators. We demonstrate how phase regularity can be integrated into the model using estimation with local models and avoiding a costly unwrapping step.

To choose among the variety of estimators available for a model, we suggest task-based quality metrics. In particular, for estimators whose output is destined to be viewed by a radiologist, we demonstrate human observer studies and models of human perception that can quantify the quality of an estimator. For features of interest that are analyzed quantitatively, we study the trade-offs between bias and variance that are available.

We find that choosing an estimator specific to the feature of interest and protocol can

produce substantially improved output. Additionally, we find that our human observer results are not predicted by SNR, challenging the use of SNR for quantifying estimator suitability. We conclude that MRI-specific estimation and evaluation provide substantial advantages over general-purpose approaches.

Keywords: magnetic resonance imaging; signal processing; image quality; point estimation; denoising

Acknowledgments

I would like to thank Dr. Burkhard Maedler, Dr. Alex MacKay, and the UBC MRI Research Centre for their assistance in acquiring data used in this thesis. Beyond providing data and time on their MRI scanner equipment, Dr. Maedler and Dr. MacKay both made themselves available to answer my questions and help direct me towards interesting applications of phase-corrected estimation in MRI. I'm also grateful to Dr. Brian Funt who made himself and his lab equipment available to assist in the development of our perception experiments.

In addition, I would like to thank my thesis supervisory committee members, Dr. Ghassan Hamarneh, Dr. Mike Hayden, and Dr. Bob Russell, who all took the time to help me with questions related to their disciplines and provided invaluable feedback on my progress throughout the duration of my studies. Outside of my committee, Dr. Richard Lockhart and Dr. Nancy Heckman both were critical in helping me formalize my ideas about MRI signal component estimation in terms of statistical theory. I must also say that the many discussions I had with my fellow members of the Medical Image Computing and Analysis lab were of great help in developing my ideas about these problems.

The involvement of my senior supervisor, Dr. Stella Atkins, has been formative of both the work presented here, and my development as a researcher. I know that I have been quite fortunate to work with a supervisor who both directed me to valuable problems and simultaneously had the patience to let me explore topics of my interest, for the sake of my own learning, even when they were not always immediately productive.

Finally, I'd like to thank Jessica Todd for her support and encouragement throughout the duration of my studies. Yes, I could probably have done this on my own, but it would have been twice as hard and half as fun.

Contents

Approval ii

Abstract iii

Acknowledgments v

Contents vi

List of Tables x

List of Figures xi

1 Introduction 1

1.1 Estimation and the MRI Signal 1

1.2 Phase-Corrected Real Reconstruction 3

1.3 Summary of Our Research Program 5

1.4 Structure of the Remaining Chapters 6

1.5 Notes on Mathematical Notation 8

2 MR Imaging 10

2.1 MRI Noise Model 11

2.2 Basics of Weighted Proton Imaging 13

2.3 Practical Issues in Weighted Proton Imaging 14

2.3.1 Physical effects that alter the signal magnitude 15

2.3.2 Physical effects that alter the signal phase 16

2.4 Solutions to Issues in Weighted Proton Imaging 17

2.5 Other Types of MRI Data 18

3 Evaluation of Image Estimators **21**
3.1 The Problem of Medical Image Evaluation 21
3.2 2-norm Error Measures 24
3.3 Task-Free Observer Experiments 26
3.4 Task-Based Observer Experiments 27
3.4.1 ROC analysis 28
3.4.2 Choosing Tasks and Observers 30
3.4.3 Comparison of task-based and task-free observer quality measures . . 32
3.5 Our Task-Based Experiments of MRI Estimator Quality 33
3.5.1 Synthetic images 34
3.5.2 Human observers 40
3.5.3 Model Observers 43
3.5.4 Human Experiment Results and Discussion 46
3.5.5 Model Results and Discussion 50
3.6 Choosing an Evaluation Method for MRI Estimators 52
3.7 Summary 55

4 Independent Estimation of Signal Magnitude **57**
4.1 Survey of Parametric Point Estimation Theory 57
4.1.1 Likelihood and maximum likelihood estimators 58
4.1.2 The Fisher information matrix and the Cramér-Rao bound 59
4.1.3 Reduction by Sufficiency 60
4.1.4 Nuisance parameters and likelihood-like functions 60
4.2 Formulation of the Estimation Problem 63
4.2.1 Two-parameter model 64
4.2.2 One-parameter model 65
4.3 Previously Published Independent Estimators 66
4.4 Derivations of Independent Two-Parameter Estimators 68
4.4.1 Maximum likelihood estimate 69
4.4.2 Maximum marginal likelihood estimate 69
4.4.3 Maximum Bartlett-corrected profile likelihood estimate 70
4.4.4 Maximum stably adjusted profile likelihood estimate 71
4.5 Comparison of Estimators 72

4.5.1 Bias Results and Discussion . . . 73
4.5.2 MSE Results and Discussion . . . 75
4.6 Conclusions About Independent Estimators for MRI . . . 78
4.7 Summary . . . 80

5 Regularized Estimation of Signal Magnitude **81**
5.1 Popular Regularization Methods . . . 81
5.2 Regularization Applied to the Magnitude Parameter . . . 83
5.3 Regularization Applied to Both Parameters . . . 84
5.4 Previous Estimators Regularizing the Phase Parameter . . . 85
5.4.1 Previous phase-corrected real estimators . . . 86
5.4.2 Other published uses of regularized phase estimators . . . 89
5.5 Derivation of Our Phase-Corrected Real Estimator . . . 91
5.5.1 The intuition behind our approach . . . 91
5.5.2 The neighbourhood model . . . 92
5.5.3 MLE of the neigbourhood model . . . 93
5.5.4 Selecting between the two maxima . . . 96
5.5.5 Choosing the neighbourhood . . . 97
5.5.6 Summary of complete algorithm . . . 98
5.6 Experiments with Real Data . . . 98
5.6.1 Fluid-filled objects . . . 99
5.6.2 Fat-suppressed knee . . . 101
5.6.3 Discussion of experimental results . . . 102
5.7 Application to T_2 Measurements . . . 105
5.7.1 Phase-corrected real estimation of T_2 data . . . 106
5.7.2 Discussion of results . . . 108
5.8 Summary . . . 110

6 Conclusions **111**
6.1 Summary of Contributions . . . 111
6.2 Future Work . . . 112
6.2.1 Task-based image estimator quality experiments . . . 113
6.2.2 Independent estimators . . . 113
6.2.3 Phase-corrected real estimation . . . 114

Bibliography 116

List of Tables

1.1 Trade-offs between estimators in four different classes. We have identified benefits (+) and detriments (−) of each estimator class. 2

List of Figures

1.1 Simulation of phase-corrected inversion recovery data. 4

2.1 Example of an MRI voxel. 12
2.2 PDF and expected value of the Rician distribution. 14
2.3 Phase measurements of GRE image with dummy objects in scanner. 19

3.1 The ideal Monte Carlo experiment for comparing medical imaging systems. . 23
3.2 ROC model of a useless imaging system. 29
3.3 ROC model of a useful imaging system. 30
3.4 Examples of feature-present synthetic images for all of the possible 18 experimental conditions in experiment one. 38
3.5 Examples of feature-present synthetic images for all of the possible 18 experimental conditions in experiment two. 39
3.6 Example of the user interface used in the human observer experiments. 41
3.7 Plot of monitor luminance vs. digital greyscale intensity used in human experiments 42
3.8 Human results in experiment one. 47
3.9 Human results in experiment two. 48
3.10 Model observer results in experiment one. 51
3.11 Model observer results in experiment two. 53

4.1 Example of independent estimator output. 74
4.2 Biases of the independent estimators. 76
4.3 MSEs of the independent estimators. 77

5.1 Real and imaginary measurements of an inversion recovery knee image. . . . 85

5.2 Example of a phase-corrected MRI voxel. 86
5.3 Demonstration of the fitting errors caused by treating angles as real values. . 88
5.4 Magnitude and phase images of fluid-filled object slice. 100
5.5 Absolute difference between magnitude and phase-corrected estimates of fluid-filled object slice. 101
5.6 Magnitude and phase images of fat-suppressed knee. 103
5.7 Comparison of Noll *et al.* at different w using fat-suppressed knee. 104
5.8 Phases of odd and even echo series in water sample. 106
5.9 Magnitude of echo series in water sample. 107
5.10 Phases of odd and even echo series in water sample after linear phase correction. 108
5.11 Phases of odd and even echo series in water sample after complete phase correction. 108
5.12 Phase-corrected estimates of s at each echo in water sample. 109

Chapter 1

Introduction

1.1 Estimation and the MRI Signal

Magnetic resonance imaging (MRI) allows us to produce images and volumes from localized measurements of magnetic properties by manipulating nuclear spin in a sample. MRI has seen wide adoption in medicine due to its ability to differentiate soft tissues which are indistinguishable in x-ray or CT images. Additionally, temporal changes in tissue properties can be used to quantify non-chemical features — functional MRI (fMRI) uses temporal variations to measure neurological activity [45] while tagged MRI can be used to measure material displacement during movement [93].

The MRI measurement, like measurements in most physical experiments, is an imperfect representation of the tissue we are imaging. In addition to noise due to electrical resistance, the recorded information can be corrupted with a wide variety of artifacts due to physical anomalies in the measurement. Since these measurements are often used as input to critical medical decisions, there is substantial demand for methods that improve the estimation of anatomical images from the MRI measurements. While it is often possible to improve the measurements by modifying the hardware or the measurement protocol, this does not avoid the question of whether a better choice of estimator could *further* improve the resulting image. From a more pragmatic perspective, having purchased a particular MRI scanner, a radiology department would like to make sure it is getting the best image possible from the significant investment in hardware.

Although we will cover the MRI signal in greater detail in chapter 2, for now it will suffice to note that MRI measurements normally assign a complex value to each discrete location

	Independent	Regularized
One-Parameter	− Ignores phase + Cannot over-smooth − Ignores image structure	− Ignores phase − May over-smooth + Uses image structure
Two-Parameter	+ Acknowledges phase + Cannot over-smooth − Ignores image structure	+ Acknowledges phase + Regularize magnitude and phase − May over-smooth + Uses image structure

Table 1.1: Trade-offs between estimators in four different classes. We have identified benefits (+) and detriments (−) of each estimator class.

(*voxel*) in the patient[1]. In the majority of experimental protocols that are used clinically, the information of interest is encoded in the true magnitude of the signal. However, since we are dealing with complex-valued measurements, the magnitude alone does not completely specify the signal; we must also describe the phase of the signal. In these protocols the true phase is the result of unknown physical processes and contains no information of value. These true magnitude and phase values are then corrupted by noise and artifacts to produce the received measurements.

Over the history of MRI a variety of estimators have been proposed that transform the complex measurements into a real valued estimates suitable for visualization as greyscale images. For the purposes of this thesis, we will classify estimators along two axes: one- or two-parameter models being the first axis, and independent or regularized estimators being the second. Table 1.1 shows some of the trade-offs between these choices. A one-parameter model uses just a magnitude parameter to describe the distribution of its measurements (normally just the magnitude measurement). A two-parameter model includes both a magnitude and phase parameter and normally describes the full complex measurement. We will see in chapter 4 that many reasonable estimators based on two-parameter models end up discarding the phase measurement. Thus, when we speak of one- or two-parameter models we are not specifying what information is used as input to the estimator, but what model is used to derive and justify the choice of the estimator function.

We use the terms independent and regularized to differentiate whether information from

[1]To those familiar with MRI this may seem an unusual description, since MRI measurements are actually complex values in the Fourier domain (k-space). We will explain the process in more detail, but here we are referring to the complex-valued voxels available after an iFFT has been applied to the k-space data.

neighbouring measurements is used to estimate individual points. We use the term independent to highlight that the distribution of the resulting estimate at any voxel is independent of the measurements at any other voxel. The same cannot be said of a regularized estimator, which takes measurements from a region around each voxel as part of its estimation at a given voxel. In asking which of these approaches is preferable we must appreciate that there are two conflicting indications: the existence of natural anatomical structure and the desire to maintain fine structures in the estimated image. By virtue of the anatomical objects imaged in clinical MRI it is reasonable to assume neighbouring voxels' underlying anatomical properties are quite similar. Anatomical structures tend to be relatively homogenous with discontinuities at their edges. An estimator that insists on independence in its estimates is essentially asserting that estimating a MRI volume with all the voxels permuted is the same task as estimating voxels in their natural order. Clearly this is false — there is information in the neighbourhood around a voxel that can improve our estimation. However, the contrary pressure is to avoid estimators that so heavily favour regularity that they remove fine structures from the estimated images. More than most other imaging applications, medical imaging focuses on the detection of small irregularities. Informally, it is probably acceptable to smooth over pimples in a family photo, but it is not acceptable to smooth over a small tumor in a medical image. The choice of how much regularization should be applied to image estimation, if it should be applied at all, is highly task dependent.

1.2 Phase-Corrected Real Reconstruction

This thesis is particularly interested with a two-parameter regularized estimation approach for MRI called *phase-corrected real reconstruction*. While we will explain how phase correction methods operate in later chapters, it is sufficient here to note that, given a two-parameter model, we can choose to regularize the magnitude, the phase, or both when we design our estimators. A phase-corrected real estimator applies regularization to the phase parameter in order to improve estimation of the signal magnitude.

MR images are normally constructed by taking the magnitude of the measurement at each location in the sample as an estimate of the true signal magnitude. Obviously this restricts the possible estimated values to the non-negative reals. Phase-corrected real images, on the other hand, can cover the whole real line. It is this ability to encode 'negative signals' that prompted the first explanation of this technique in a general form [2]. In particular,

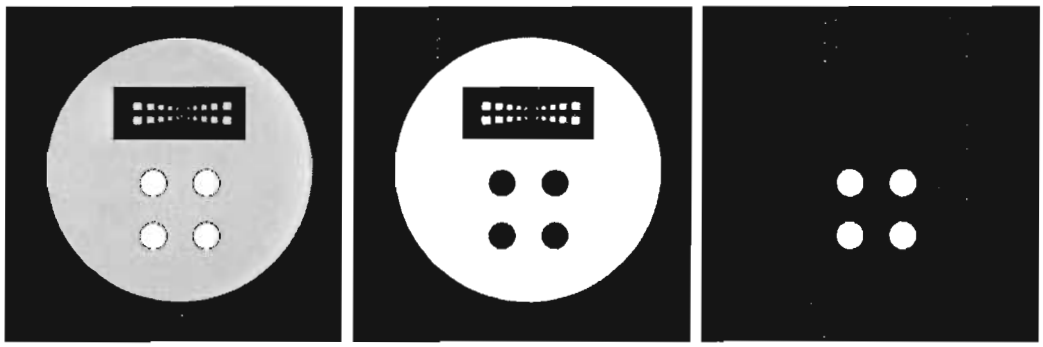

Figure 1.1: Simulation of the effect of phase-correction on inversion recovery data. All three images depict a simulated system with a large cylinder of water containing an empty region of air with small water-filled structures (top of cylinder) and four smaller cylinders of oil producing slightly greater signal magnitude (bottom of cylinder). The water is simulated as producing a signal with positive sign while the oil's signal is given a negative sign. *Left*: Inversion recovery image produced using the standard magnitude reconstruction. *Center*: Positive part of inversion recovery image produced by phase correction. *Right*: Negative part of inversion recovery image produced by phase correction. Note that the fat (four small cylinders) and water (main body) are separated into the negative and positive images after phase correction while the magnitude reconstruction discards the sign and thus blends the two signals.

it was noted that inversion recovery MR images, where two signal sources (usually fat and water in medical MRI) can be assigned opposite signs, were far easier to interpret when the images maintained the signs of the signal source . Figure 1.1 shows an example of this, where the magnitude reconstructed inversion recovery data (left image) cannot distinguish water and fat while the positive part of the phase-corrected real reconstruction inversion recovery data (center image) contains only water and the negative part (right image) contains only fat.

Interest in phase-corrected real reconstructions was rekindled by a theoretical work showing that magnitude images would have less contrast than phase-corrected images [13] (we will revisit this result in chapter 3). There have since been a succession of proposals for performing phase-corrected real reconstructions [14, 49, 58]. However, several authors have published concerns about the robustness and significant computational demands of these methods [49, 59] and they are not routinely used as part of clinical MR imaging.

In chapter 5, we suggest that one reason previous phase correction methods have tended

to be unusable in practice is that they focused on simultaneous phase unwrapping and phase-corrected real reconstruction. Phase unwrapping problems appear in many application areas outside of MRI and are generally quite difficult to solve efficiently. The basic problem takes as input an array whose entries are angles and asks for a solution which consists of the smoothest array of real numbers such that the solution equals the input modulo 2π. The difficulty in this problem varies by the exact definition of 'smoothest' being employed, but in general this problem often reduces to some form of multivariate non-linear optimization. We can, however, perform phase-corrected real reconstruction without worrying about unwrapping — this thesis demonstrates one such algorithm and suggests a methodology for constructing other algorithms based on the same principle.

1.3 Summary of Our Research Program

We were interested in whether an algorithm for phase-corrected real MR images could be produced that would be relatively fast, robust given clinical data, and actually improve MR image quality. We were particularly interested in applications of these reconstruction approaches to MRI measurements that have quite low signal-to-noise ratio (SNR). In this thesis we take concrete steps towards answering all of these questions in the affirmative.

Interestingly, of the topics we cover in this thesis, concerns about whether phase-corrected real MRI could improve MR image quality were the ones least addressed in previous publications. This is problematic: if we are not sure that phase-corrected real reconstructions really improve MR image quality, why should researchers be pursuing this reconstruction approach? Our reading showed only a theoretical result on this topic (Bernstein *et al.* [13]) but no indication of what effect this result might have in practice.

At first blush it might seem that this is a chicken-or-egg problem. If there are no phase-corrected real reconstruction algorithms to test, how can we perform experiments to determine whether phase correction is an improvement over magnitude? Combining previous work in measuring image quality in other modalities with a general model of the MRI signal, we developed a series of experiments using real human observers to test the *best-case* outcome from phase-corrected real reconstruction against alternative reconstructions using synthetic images. Our results suggest that the best-case phase-corrected real reconstruction could indeed provide an improvement over other estimators when searching for faint targets in dark parts of the background. While we found that some of our results for other estimators

contradicted previous theoretical assertions, this result for the phase-corrected real estimator correlates with the enhancement expected by the previous theoretical result [13].

Having experimental indication that phase-corrected real reconstructions could improve detection, we were still left with the problem that none of the existing algorithms for phase correction were considered reliable for clinical use due to the difficulties of regularizing the phase parameter. Since regularization was the chief difficulty, we were curious whether an independent estimator constructed from the two-parameter model could approximate the best-case phase-corrected real estimator's performance without requiring regularization. In particular, we were interested in applying the large body of statistical research on estimation in models with nuisance parameters to the two-parameter model of the MRI signal. The great majority of previous work on independent estimators has focused on one-parameter models, but we were interested in whether one of the existing approaches in the statistical literature might produce a result similar to phase-corrected real estimators without requiring regularization. In pursuing this idea we developed a variety of independent two-parameter estimators and quantified their behaviour in terms of bias and mean squared error (MSE). While some of these new estimators proved competitive with state-of-the-art independent estimators, none of them provided performance similar to the ideal phase-corrected real reconstruction.

Having determined that regularizing phase in a two-parameter model does indeed provide advantages over the alternative estimators, we derived a new algorithm for phase-corrected real reconstruction. Our estimator is different from many previously published algorithms in that it is derived with the same method we use to derive our two-parameter independent estimators. Additionally, our derivation focuses on the problem of signal estimation and avoids unnecessary work such as phase unwrapping. Comparisons between our approach and previous phase estimation algorithms have demonstrated promising results in a variety of real and synthetic data sets.

1.4 Structure of the Remaining Chapters

The majority of work in this thesis rests upon a model of the MRI signal originally presented by Henkelman [34]. It is this model that makes up the basis of what we refer to as the two-parameter model for the MRI signal. In chapter 2 we introduce and explain the Henkelman model in detail, along with the most oft-encountered exceptions in practice. We also show

how the model with magnitude and phase parameters describing the signal is reduced to a model with just a magnitude parameter via marginalization.

In chapter 3 we survey the various approaches to measuring medical image quality that have been used in previous work. We then explain how we used our model of the MRI signal, along with the previous task-based experimental methodologies in other medical imaging modalities, to develop new experiments for the evaluation of MRI estimators. We present the results of our experiments and draw the conclusions that motivate the remainder of this thesis.

Chapter 4 begins with a brief summary of the estimation theory used to derive independent estimators from the two-parameter model of MRI. In particular, we focus on the frequentist concept of point estimation from likelihood and likelihood-like functions. We discuss some of the theoretical approaches presented in the statistics literature for estimation of a single parameter in a multi-parameter model. We then apply these techniques to the models derived from MRI and derive several new estimators for the MRI signal magnitude. We compare these new estimators with previously published estimators and the results expected from the ideal phase-corrected real estimator.

Our proposal for a novel phase-corrected real estimator is presented in chapter 5. The chapter begins with a survey of previous phase correction approaches and other phase estimation algorithms that could be used as part of a phase correction algorithm. We explore the difference between phase unwrapping and phase-corrected real estimation and highlight how the former is unnecessary for the latter. Based on this differentiation we present a slight modification of our two-parameter model which, by applying the standard estimation theory outlined in chapter 4, naturally produces phase-corrected real estimators. We compare the most basic of these estimators to previously published algorithms for phase-corrected real reconstruction using real and synthetic data.

Finally, in chapter 6 we summarize the results of the previous chapters and suggest some avenues for future research that appear promising based on the results of this thesis.

Please note that, because of the multi-part (and multi-disciplinary) nature of our research program, this thesis uses a slightly different structure than is standard. Where there would normally be one or more stand-alone chapters to cover previous work, here we have included the relevant previous work at the beginning of each chapter. This structure hopefully highlights exactly what previous work applies to each part of our research program while allowing the chapters to represent logical parcels of the research program.

1.5 Notes on Mathematical Notation

Before proceeding it is useful to specify some very general rules for mathematical notation that are applied for the rest of the work. Throughout the thesis, notation is defined as it is introduced and used. However, some very general rules can improve the clarity of the equations. Since we must choose some convention among the many available, we have followed the ISO guidelines for engineering mathematics typography [12].

- We will use lowercase italic type like x for complex variables, lowercase bold italic type like $\boldsymbol{x}$ for vector variables, and uppercase bold type like $\mathbf{X}$ for matrices.
- We will use lowercase upright type like i for constants such as π, i and e.
- We will use $\{\,\}$, $[\,]$ and $(\,)$ as necessary to clarify nested parentheses. The choice of bracket generally has no special meaning. The one exception to this is that when we are referring to an element of a matrix $\mathbf{M}$ that is addressed by a 2-vector index $\boldsymbol{x}$ we will use square brackets and write $\mathbf{M}[\boldsymbol{x}]$ for the element.
- $\|\boldsymbol{x}\|_p$ is the p-norm of the vector $\boldsymbol{x}$. When no p is specified, $p = 2$ is implied.
- $P(x > y; \alpha)$ is the probability of the measurement x being greater than y with parameter α fixed.
- $p(x; \alpha)$ is the probability density function (PDF) for the measurement x with parameter α fixed.
- $L(\alpha; x)$ is the likelihood of the parameter α with measurement x fixed.
- $\ell(\alpha; x)$ is the log-likelihood of the parameter α with measurement x fixed.
- $U(\alpha; x)$ is the score of the parameter α with measurement x fixed.
- $E(x)$ is the expected value of x.
- $\hat{x}$ is an estimate of x.
- $\mathcal{N}(\mu, \sigma)$ is the normal distribution with mean μ and standard deviation σ.
- $\frac{\mathrm{d}\boldsymbol{x}}{\mathrm{d}\boldsymbol{v}}$ is the matrix of total derivatives of each element of vector $\boldsymbol{x}$ with respect to each element of vector $\boldsymbol{v}$. This generalizes to scalars by treating them as 1-vectors.

- $\frac{\partial \boldsymbol{x}}{\partial \boldsymbol{v}}$ is the Jacobian matrix of the vector $\boldsymbol{x}$ with respect to the vector $\boldsymbol{v}$. This generalizes to scalars by treating them as 1-vectors.
- $\mathbb{R}^+$ is the set of non-negative reals.
- $|\mathcal{S}|$ is the cardinality of set $\mathcal{S}$.
- $\mathrm{I_n}(z)$ is the n^{th}-order modified Bessel function of the first kind with argument z.

Chapter 2

MR Imaging

MRI is most commonly used to record volume data sets from human or animal subjects. In a setup with a single quadrature receiver coil, the measurements are normally acquired in the complex-valued Fourier domain [16], which is referred to as k-space in the MRI literature with the individual frequency axes being labeled k_x, k_y, k_z. In some cases the z-axis is sliced during acquisition, resulting in a stack of 2D Fourier slices instead of one 3D Fourier volume. Since the k-space measurements are usually acquired on, or re-gridded to, a Cartesian grid, the k-space measurements can be efficiently transformed into a complex-valued image volume via the inverse fast Fourier transform (iFFT). In the stack-of-2D-slices acquisition approach, the iFFT is applied to each slice individually, while in the 3D-volume acquisition a 3D iFFT is applied. Since our interest will mostly be on the spatial-domain measurements, in the remainder of this thesis when we refer to "measurement" without specifying whether it is pre- or post-iFFT, we will take it to mean the post-iFFT spatial measurements.

The process of acquiring the k-space measurements relies on the interaction of magnetic fields with, most commonly, the protons associated with hydrogen atoms in the subject's body. The exact timing and amplitude of the magnetic fields and the timing of the measurement of voltage across the receiver coil (termed *signal readout*) determines the details of the received signal's relationship to other physical properties of the tissue. A particular configuration of magnetic field application and signal readouts is called a *pulse sequence*. Although the received signal is generally proportional to the density of protons at a given location, the selection of a particular pulse sequence can have a substantial impact on the

final image. For example, a delay between the excitation of the protons and their measurement will lead to tissues with fast spin-spin relaxation (also called T_2 relaxation) appearing darker in the image than tissues with slower relaxation. A variety of good textbooks are available that cover the details of the physics of pulse sequences (e.g., [85, 40, 44]) and so for brevity we will not include a survey of this topic here. Instead we will focus on a higher-level model of the recorded measurements that explains the properties of interest without delving much into the underlying physics.

2.1 MRI Noise Model

In the majority of clinical pulse sequences the information of interest is encoded as the true magnitude of each complex-valued voxel. Thus, it is very common to display a greyscale image of the voxel magnitudes as the image a radiologist will use for diagnosis. One question that naturally arises is how the noise of the physical system is distributed in these images. The first coherent description of the noise in MR measurements and the resulting distribution in magnitude images was given by Henkelman [34]. The basic form of this model comes from the assumption that MRI measurements recorded using a single quadrature coil are corrupted at every voxel with two independent samples from the same zero-mean normal distribution. One sample is added to the real channel and one to the imaginary channel. Further, since the noise results from random thermal fluctuations associated with the effective resistance of the patient combined with that of the coils themselves, the variance of the normal distribution is the same at all locations in k-space [35, 27, 48]. Fortunately, the noise in this model is additive. Thus, the iFFT of these measurements is the same as the iFFT of the true, noise-free complex signal summed with the iFFT of the noise. The noise in the two channels is independent and drawn from $\mathcal{N}(0, \sigma)$, which implies the noise has the same distribution in the spatial domain, although the individual noise samples will be different [16]. Combining these results, we find that our complex-valued spatial-domain measurements are the true, noise free complex-valued signals summed with a complex zero-mean normally distributed noise.

While Henkelman proceeded from this model to derive the distribution of the magnitude image, he did not solve the integrals in his equation nor note its relationship to previous work in signal processing. This final step was taken by Bernstein, Thomasson, and Perman [13] who recognized the magnitude distribution as the Rician distribution [66] and thus

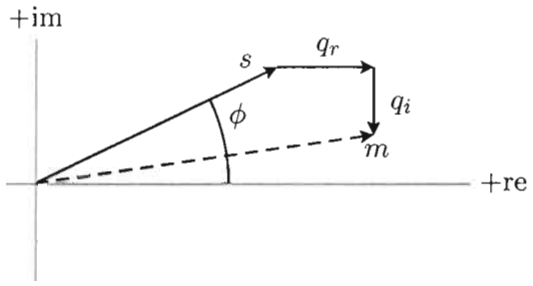

Figure 2.1: Example of an MRI voxel, showing the relationship of the various components in equation (2.1). The solid black lines represent the three components of our signal model and the indicated angle is the signal phase. The dashed line shows the sum of the components. This dashed line is the recorded value of the voxel while all the other parts of the diagram are unknown.

enabled an integral-free description of the magnitude image's distribution. This was further expanded by Gudbjartsson and Patz [32]. Recently there has been doubt raised about whether the assumption that the two channels are uncorrelated truly holds for MR imaging [30]. However, while it is likely that using a more complicated model might further improve the results of our work, we will focus on the work descending from Henkelman's model as this has been the source of the most results up to the present. Additionally, the methodologies we will describe in later chapters can all be extended to other signal models as well.

In all of the problems discussed in this work we will consider pulse sequences that encode the clinically relevant information in the magnitude. However, in order to understand the probabilistic behaviour of the measured magnitude at the receiver, we need a model of the noise in the complex-valued MRI measurement. If we consider the measured value m at a single complex-valued voxel, we can write m as the sum of components

$$m = s\exp(\mathrm{i}\phi) + q_r + \mathrm{i}q_i\,. \tag{2.1}$$

where $s \in \mathbb{R}^+$ is the magnitude of the true signal produced by the anatomy, $\phi \in (-\pi, \pi]$ is the phase of the true signal, and q_r and q_i are the noise in the real and imaginary components respectively. A single voxel produced by this system is depicted in figure 2.1 to illustrate how the components interact.

Since we have stated that q_r and q_i are independent samples at every voxel, we can write the PDF of the complex-valued measurement at a given voxel as

$$p(a, b; s, \phi, \sigma) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(a - s\cos(\phi))^2 + (b - s\sin(\phi))^2}{2\sigma^2}\right), \tag{2.2}$$

where a and b are the real and imaginary parts of m, and σ is the noise standard deviation. Alternately, we can convert the measured value to polar coordinates and write the PDF as

$$p(r, \theta; s, \phi, \sigma) = \frac{r}{2\pi\sigma^2} \exp\left(-\frac{s^2 + r^2 - 2sr\cos(\theta - \phi)}{2\sigma^2}\right), \tag{2.3}$$

where r and θ are the magnitude and phase of m.

The principal virtue of this model, particularly for the work we describe, is that it is not concerned with what information is encoded in s and ϕ and thus works for almost any single-coil pulse sequence. When we consider specific pulse sequences, we can expand the s and ϕ variables as necessary based on our model of the pulse sequences' physics.

2.2 Basics of Weighted Proton Imaging

The most common form of MR image used clinically is the weighted proton image. In these images, the magnitude of the true, noise-free signal at each voxel represents the number of protons available for imaging in the voxel volume weighted by some physical property of the tissue in the voxel. Common properties that are used in weight are spin-lattice relaxation (T_1-weighted), spin-spin relaxation (T_2-weighted), and diffusion (DWI). An image that is not weighted by any tissue property is called a proton density (PD) image since its true signal magnitude reflects only the number of protons available for imaging in the tissue.

In principle, all of the quantities being estimated in weighted proton imaging are non-negative real values. The pulse sequences that produce these images encode the information as the signal magnitude, s, leaving ϕ as a free variable. Spin echo (SE) pulse sequences attempt to set $\phi = 0$ for all voxels, while gradient echo (GRE) pulse sequences do not concern themselves with the value of ϕ and expect it to vary from voxel to voxel. For now we may treat this variation as random, but later we will consider whether these variations have some underlying structure.

Images are usually produced from these measurements by taking the magnitude at every point as the estimate of s. As noted above, this gives an image with a Rician noise distribution. To see this, we can marginalize out the phase measurement from equation (2.3) to produce the Rician PDF

$$\begin{aligned} p(r; s, \sigma) &= \int_{-\pi}^{\pi} p(r, \theta; s, \phi, \sigma)\, \mathrm{d}\theta \\ &= \frac{r}{\sigma^2} \exp\left(-\frac{r^2 + s^2}{2\sigma^2}\right) \mathrm{I}_0\left(\frac{sr}{\sigma^2}\right), \end{aligned} \tag{2.4}$$

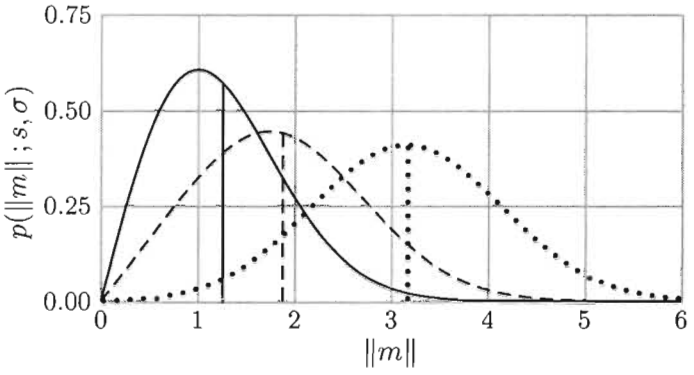

Figure 2.2: PDF and expected value of the Rician distribution. The x axis is the observed magnitude $\|m\|$ while the y axis is the value of the PDF for the appropriate combination of parameters. The distribution is produced as the magnitude of a complex value consisting of the sum of a real-valued signal s and two samples from $\mathcal{N}(0,\sigma)$. The curves plot the PDFs of distributions with $\sigma = 1$ and s being set to: 0 (solid line), 1.5 (dashed line) and 3 (dotted line). The vertical solid, dashed, and dotted lines joining the distribution to the x axis demarcate the expected value of the distribution in the three values of s.

using the identity $\int_{-\pi}^{\pi} \exp(z\cos(\theta))\,\mathrm{d}\phi = 2\pi\,\mathrm{I}_0(z)$, where $\mathrm{I}_0(z)$ is the zeroth-order modified Bessel function of the first kind. One feature of the Rician distribution that has been much-discussed in the MRI literature is its bias when s is small relative to σ (see figure 2.2). Because the bias varies with s, it has been suggested that regions of the image with low s relative to σ may have reduced contrast [34, 13, 32]. Consulting figure 2.2, we can see that the bias is quite large when $s/\sigma = 0$, but diminishes quickly, with the noise distribution becoming approximately normal for $s/\sigma \geq 3$.

Using the model in equation (2.1) we can also now express the basic idea of phase-corrected real reconstruction: if ϕ is known or can be estimated, we can multiply m by $\exp(-\mathrm{i}\phi)$, thus aligning s along the real axis. We can then take just the real component of the phase-corrected measurement as the estimate of s, which will contain s summed with a zero-mean normally distributed noise.

2.3 Practical Issues in Weighted Proton Imaging

The probabilistic model described in the previous section describes the effects of electrical resistance on the measurements. However, a variety of physical effects also alter the parameters of these distributions. We will term these alterations *artifacts* to differentiate them

from the random noise. We will think of noise as the difference between the measurements and the distribution parameters, while artifacts alter the parameters from their ideal values. This section provides examples of some of the most common artifacts visible in MRI measurements.

2.3.1 Physical effects that alter the signal magnitude

Many physical phenomena can cause the magnitude of the true signal s at a particular voxel to include more than just the weighted proton density at that location. One of the most visibly apparent is caused by movement during the scan. Movement of a patient's tissue or the flow of blood during imaging can produce a 'ghosting' effect where the anatomy appears faintly in multiple locations. Since the time required to acquire a single k_y row of k-space is quite short in most pulse sequences, it is commonly assumed that each row is recorded with the patient in a fixed location. However, inter-acquisition motion (also called inter-view motion) results in the k_y rows acquired before and after the movement being inconsistent with any one patient location. This inconsistency produces the image ghosts when the inverse Fourier transform is applied [33].

In addition to motion artifacts caused by blood flow, many MRI pulse sequences suffer from 'wash in' and 'wash out' effects when they image slice-by-slice. Blood that was not excited 'washes in' and produces a signal void in the slice. In other cases the excited blood vacates the slice producing a signal void in one location and an undesired 'hot spot' in a different part of the body. Further, flowing blood can suffer from motion artifacts that cause the blood to appear outside the vessel that is traveling inside [85, 33]. A variety of modifications to standard pulse sequences have been proposed to reduce these effects.

Chemical shift is the name applied to the imaging errors due to the difference between the magnetic fields experienced by the hydrogen protons in water and fat molecules (for more on chemical shifts and their relationship to pulse sequences see [85, 40, 44]). In a water molecule, the hydrogens' electrons are strongly attracted to the oxygen atom. In fat, we have hydrogen-carbon bonds with less pull on the hydrogens' electrons. These differences in bonding result in each proton experiencing a different magnetic field depending on the tissue type. The principal effect from chemical shift in most pulse sequences is that the fat and water signal will be shifted relative to each other in particular directions called the *frequency encoded* directions of the pulse sequence (usually k_x by convention). The magnitude of this shift is influenced by the choice of pulse sequence and, with appropriate

choice of parameters, can often be reduced to less than one voxel, effectively eliminating the shift [33].

A long-standing difficulty for all MRI systems has been the production of the large, time-invariant, homogenous magnetic field needed for proper imaging. A large amount of the mathematics of MRI relies on precisely controlling the magnetic field at every point in the scanner. A major source of problematic variations are the fields induced by each substance when exposed to the main field [69, 85]. Each substance has its own magnetic susceptibility and when two regions immediately adjoining each other have different susceptibilities, the total magnetic field changes rapidly along the regions' boundary. This effect is often noticed at tissue-air boundaries in regions like the sinuses. Calculating the precise shape of these induced fields is often difficult even when the precise shape of the objects is known [69]. In general, the more curved the boundary, the larger the magnetic field gradient induced along the boundary [85].

Significant changes in field strength over distances on the order of a micron have an effect on the T_2 value of the sample. This is because the water molecules, and thus protons, being imaged tend to diffuse at about $1.2 \times 10^{-5}\text{cm}^2/\text{s}$ in human tissue. This diffusion, combined with a spatially varying magnetic field, causes protons to experience many field strengths during the pulse sequence [69]. For this reason, magnetic field variations tend to cause the observed T_2 relaxation to be faster than the tissue's inherent T_2. To account for this, a different, unknown constant T_2^* is often used to describe relaxation instead of T_2, with the understanding that $T_2 > T_2^*$. In practice, this means that we will experience signal loss in areas with large inhomogeneity over short distances.

When the main magnetic field has a more gradual spatial variation, the resulting effect appears more like chemical shift. Areas where the field is stronger will be shifted in the positive x direction while those experiencing a weaker field will be shifted in the negative x direction or vice versa depending on the pulse sequence used.

2.3.2 Physical effects that alter the signal phase

We previously stated that GRE imaging ignores the phase ϕ while SE attempts to force $\phi = 0$. However, we did not go into detail about what causes ϕ to take any particular value. When imaging a perfectly homogenous object with a perfectly homogenous magnetic field, we expect $\phi = 0$ to result from either of the pulse sequences. In practice, deviations from this perfect model may cause ϕ to take other values.

When imaging humans it is obvious that the sample is not homogenous. We expect to encounter both chemical shift and magnetic field inhomogeneity as described in our discussion of magnitude artifacts. Both of these effects will cause the value of ϕ in GRE images to vary, advancing where the field is stronger and lagging where the field is weaker. However, because the phase is an angle, and thus wraps around the circle, we cannot make simple greater-than/less-than statements about the resulting phases.

SE imaging is able to correct for chemical shift and magnetic field inhomogeneity effects that are constant for the duration of the scan. However, if the field changes slightly during the scan, either due to patient motion, temperature changes in a fixed magnet, or new fields induced by the magnetic field switching required by the pulse sequence, then we will still find that $\phi \neq 0$ in SE signals. In general, we expect the amount of phase advance or lag experienced by any point in an SE image to be substantially less than the equivalent point imaged using GRE.

2.4 Published Solutions to Practical Issues in Weighted Proton Imaging

Due to the long acquisition times required for MR imaging, correction of patient motion is an important area of research. In chest imaging, periodic cardiac and respiratory motion cannot be voluntarily paused for the duration of an imaging sequence. In these cases 'gating' is often used so that measurements are always collected at the same point in the biological cycle. It is hoped that the anatomy will be in approximately the same position at this point in each cycle, so the k_y rows will represent a consistent state. Another solution makes use of the fact that very often the anatomy being imaged can be treated as a rigid body. For example, the head and brain tend to move as one rigid unit only undergoing rotations and translations. Many different proposals have been made for compensating for rigid body motion in an MRI slice. Approaches for detecting these motions and compensating for them either during the scan (e.g., [78, 83]) or via post-processing (e.g., [95]) are a significant area of research.

To correct deviations in the main field that exist for a long period of time (e.g., variation due to the magnet's construction, or due to external influences at the installation site) smaller *shimming* magnetic fields are frequently applied to produce the desired homogeneity. However, in practice inhomogeneities that change over time (e.g., inhomogeneities that

move with the patient) cannot be effectively counterbalanced with shimming, although some approaches for real-time correction of shimming errors introduced by patient motion have been proposed recently (e.g., [83]).

Finally, a variety of algorithms for turning the received measurements into estimates of signal magnitude have been published. Most these approaches focus on compensating for the measurement noise described in equations (2.3) and (2.4), and do not account for artifacts that alter the true signal magnitude. We will continue with this focus, leaving artifacts that alter signal magnitude as a separate issue. Using the classifications presented in chapter 1 we can separate the previous approaches to this problem. Although our work focuses almost exclusively on two-parameter estimator, previous work using the one-parameter model to generate independent estimators for this problem will be surveyed in chapter 4 and previous one-parameter regularized estimators will be presented in chapter 5.

While both our work and almost all of the previously published works on estimators ignore alterations of the true signal magnitude, the causes of variation in the true phase ϕ are important to phase-corrected real reconstruction and so previous authors have paid substantial attention to them. The most important fact about ϕ for these applications is that, in the majority of pulse sequences, the artifacts that affect the true phase vary quite smoothly across the spatial extent of the image. Figure 2.3 shows the phase map for a weighted proton image of water and cream containers, demonstrating the smooth phase variation common in MRI. In particular, many previous authors have focused on fitting low-degree polynomials to ϕ with relative success [2, 14, 46, 19]. It is important to note that in most pulse sequences these smooth variations occur only within MR image slices and we should not expect ϕ to vary smoothly across slices. However, it still allows strong assumptions to be made about ϕ, and we will rely on this significant result when we derive our two-parameter regularized estimator in chapter 5.

2.5 Other Types of MRI Data

Since the majority of our work will focus on weighted proton imaging, we will not discuss other types of MRI data in great detail. However, two other types of MR data require description as we will refer to them later in the thesis.

Inversion recovery data is similar to weighted proton imaging in that the information is encoded in s. However, as we noted in chapter 1, inversion recovery images differ in that

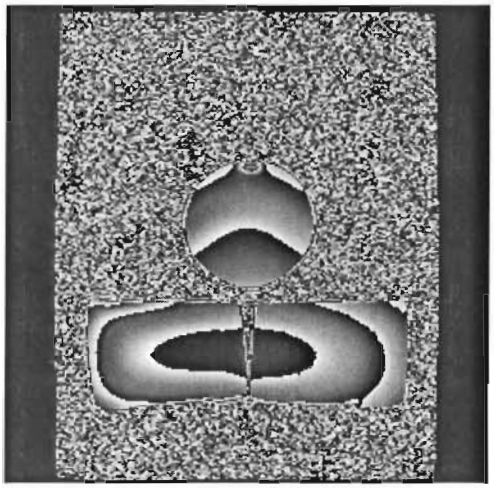

Figure 2.3: Phase measurements of GRE image with dummy objects in scanner. Phase is mapped to greyscale with $-\pi$ being black and π being white. However, phases of $-\pi$ and π represent the same orientation, since they are the same point when wrapped onto the circle of angles. Thus, black/white edges are an artifact of the color map, and actually represent smooth variations in the phase measurements. The speckle outside the objects is the result of the measurements containing pure noise, and thus the voxels have incoherent phase.

we allow s to take on negative values. If we allow this change to the definition of s, then equation (2.3) can still be used to describe the PDF of the measurements. Similarly, all of the physical effects that we described for weighted proton images apply to these data as well. Magnitude-reconstructed inversion recovery images suffer from Rician bias, but also from the loss of sign information depicted in figure 1.1.

T_2 echo series are used to perform quantitative analysis of the T_2 relaxation at each location in the subject. We have previously stated that by selecting an appropriate pulse sequence we can weight the proton image with a function that depends on T_2. In the usual method of T_2-weighted imaging, the weighting function is

$$s = s_0 \exp[-t/T_2]\,, \tag{2.5}$$

where s_0 is the unknown true signal magnitude that would be observed without any weighting, t is a parameter that we may choose, and T_2 is the unknown relaxation constant of the voxel being imaged.

In general one of these T_2-weighted images alone cannot be used for quantitative estimation of T_2 since it is impossible to distinguish variations in s_0 from variations in T_2. However, since we know the form of the weighting function and can control the value of t for any given image, we can estimate the T_2 properties if we have a series of images with different t values. In experiments of this type it is normal to collect many such images for a sequence of increasing t values. Due to the underlying physics of how the signal is acquired, each measurement in the sequence is referred to as an *echo*. As we can see in equation (2.5), the T_2-weighting is a dampening effect, so later echoes will have smaller s than earlier ones. Additionally, applying enough T_2 weighting will effectively completely null the signal at a given voxel. However, we have seen that even in regions with no signal there is still a positive bias due to Rician noise if we take the magnitude of each echo as our measurement. The combination of gradual exponential decay with the Rician-distributed measurements means that the bias of the measurements will increase as the decay increases. If this variation in bias is not accounted for it can distort quantitative analysis of T_2 measurements. We will discuss this problem, and present a possible solution, in chapter 5.

Chapter 3

Evaluation of Image Estimators

The first component of our research program was to determine if phase-corrected real reconstructions could improve MR image quality. Although there has been previous experimental work on MR image quality [68, 54], to the best of our knowledge there has only been one theoretical result on the potential advantages of phase-corrected real MRI [13]. We begin this chapter with an introduction to the basic problems inherent in quantifying medical image quality. This is followed with a survey of some of the approaches that have previously been used for measuring medical image quality. We then describe two experiments we conducted with human participants to quantify the quality difference between the output of three different MRI signal magnitude estimators and discuss the results. We use these results to justify our decision to pursue the development of estimators that perform like the best-case phase-corrected real estimator. Returning to our survey of evaluation approaches, we provide an explanation of how we will quantify and compare estimators for the remainder of this thesis and relate our choice to our experiment results. We finish the chapter with a brief summary.

3.1 The Problem of Medical Image Evaluation

Of the many ways we could evaluate an estimator, the most common is to quantify its fidelity to an ideal estimator. In the common signal processing model, a signal is communicated by a source, via a channel, to a receiver. The channel modifies the signal (e.g., adds noise, performs some transformations, etc.) so that the original signal does not arrive at the receiver. An ideal estimator in this model is one that can, given the measurements made

at the receiver, reconstruct the original signal. This suggests that we should evaluate our image estimators based on their output's fidelity to the original image.

The critical issue that arises with this approach is the need to define what we mean by "fidelity". While, in theory, any measure will suffice, in practice it is likely that certain measures are more meaningful for a given estimation problem. In particular, since we often model our channel as a random process, we must equally view our estimator as a random function of the signal. Thus, our problem is not just one of summarizing the difference between two estimates, but summarizing the difference between the ideal and a random function.

The challenge of evaluating an estimator in the context of medical imaging is often further compounded by the fact that the image is only a part of a larger process involving a radiologist. We can view the patient's state as the source, the imaging process as the channel, and the radiologist's report as the estimate based on the received information. In this case any estimated images we produce are viewed by radiologists as part of their communications channel with the patients. However, since we cannot modify the radiologists, we experiment with different proposed image estimators and compare how they affect the fidelity of the radiologists' output to a gold standard experiment measuring the patients' disease states.

One approach that is often considered ideal for the medical imaging context involves constructing a Monte Carlo experiment using a complete clinical environment. To do this, we compare the quality of medical imaging systems (e.g., compare estimators, compare hardware, compare displays, etc.) by having a set of radiologists generate reports about a set of subjects using each imaging system. We would then compare these reports with some gold standard experiment (e.g., biopsy) to measure the error rate of the system. Comparing the error rates, we claim the imaging system which produced reports more consistent with the ground truth was the superior system for the task under study. This 'ideal' experimental methodology is depicted in figure 3.1.

While empirically satisfying, this sort of experiment is prohibitively expensive to run, requiring a significant amount of radiologist time reading experimental images and additionally a substantial collection of clinically useless patient data be captured using experimental systems unsuitable for diagnosis. In practice, it is necessary that we be able to evaluate novel estimators for MRI without such an elaborate experiment. Ideally, as researchers, we would like algorithms that allow us to compute the quality of image estimators without recourse to the radiologist. Obviously, new methods destined for clinical practice need a

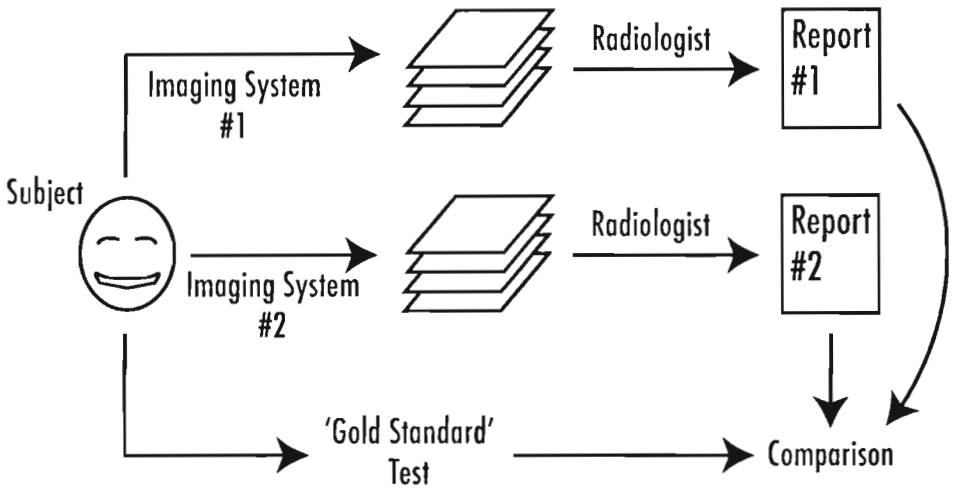

Figure 3.1: The ideal Monte Carlo experiment for comparing medical imaging systems uses information from subjects (represented with the face on the left) collected via three information pathways: the two imaging systems under comparison and a 'gold-standard' test such as biopsy. The imaging systems each generate image stacks that are used by different radiologists to generate independent reports. By evaluating reports generated using the two imaging systems against the results of the 'gold-standard' test, we can determine which imaging system is superior for a given task.

more thorough evaluation, but we are interested in first-line methods of validation suitable for judging research proposals. The next three sections outline three different approaches to performing these sorts of research experiments.

3.2 Quantifying Quality with 2-norm Error Measures

Error measures attempt to quantify how similar an estimate or estimator is to the true signal. To evaluate the error of an estimate, we are measuring how similar it is to the true signal. This is distinguished from the error of an estimator which must summarize the estimator's behaviour given the random nature of the communications channel. The error ascribed to an estimator will, in turn, rely on how we determine the error of any estimate it produces.

The terms *mean squared error* (MSE) , *root mean squared error* (RMSE), and *root mean squared deviation* (RMSD) are all commonly used to describe a choice of error measure for an estimate and an estimator. The basis of all of these measures is that the 2-norm of a vector emphasizes a few large entries more than a lot of small ones. Thus, given two images represented as vectors $\boldsymbol{x}$ and $\boldsymbol{y}$ of length n, we can see that minimizing $\|\boldsymbol{x} - \boldsymbol{y}\|$ will tend to encourage spreading the differences between the two vectors over as many entries as possible and discourage a few entries with large differences. This describes one intuition about what it means for vectors to be similar. However, the exact definitions and meaning of MSE, RMSE, and RMSD often vary slightly between disciplines. For the sake of clarity, we will define what we mean when using them.

We define RMSD for two n-vectors $\boldsymbol{v}$ and $\boldsymbol{u}$ as

$$\mathrm{RMSD}(\boldsymbol{v}, \boldsymbol{u}) = \frac{\|\boldsymbol{v} - \boldsymbol{u}\|}{\sqrt{n}} . \tag{3.1}$$

Moving from vectors to a real estimator $\hat{x}$ of a real quantity x, the MSE and RMSE are defined as

$$\mathrm{MSE}(\hat{x}) = E\left[(\hat{x} - x)^2\right] , \tag{3.2}$$

and

$$\mathrm{RMSE}(\hat{x}) = \sqrt{\mathrm{MSE}(\hat{x})} . \tag{3.3}$$

Thus, if we take n samples from estimator $\hat{x}$, we can use the RMSD between the n-vector of samples and an n-vector containing x at every entry as an estimate of the RMSE and

square this to get an estimate of the MSE. This is the Monte Carlo estimate of MSE that we will return to in chapter 4.

Two other common quantities that arise in the evaluation of real-valued estimators are *bias* and *variance*. These are defined for an estimator $\hat{x}$ as

$$\text{Bias}(\hat{x}) = E(\hat{x} - x)\,, \tag{3.4}$$

and

$$\text{Var}(\hat{x}) = E\left[(\hat{x} - E(\hat{x}))^2\right]\,. \tag{3.5}$$

These can be related to MSE by the formula

$$\text{MSE}(\hat{x}) = \text{Var}(\hat{x}) + (\text{Bias}(\hat{x}))^2\,. \tag{3.6}$$

The application of these measures to imaging is where the definitions become important. For example, if we have a greyscale image $\boldsymbol{v}$, a channel that independently adds noise to each pixel, and an estimated greyscale image $\boldsymbol{u}$, we can use RMSD to measure the difference between the images. However, it is well known that a high RMSD does not correlate well with there being a noticeable difference in the images (it is often stated that high MSE does not correlate with perceptible difference, but this is due to RMSE being used to describe what we have called the RMSD)[29]. For example rotating all the columns ahead one pixel such that the last column on the right becomes the first column on the left will generally produce a high RMSD, but perceptually the only difference will be a small band on the left of the image.

The problems with RMSD as an image quality measure might lead one to question whether the 2-norm should be used at all in evaluating estimators for imaging. Let $\hat{\boldsymbol{v}}$ be a vector-valued estimator of the vector $\boldsymbol{v}$. It is important to note that, given the definitions above, the MSE of $\hat{\boldsymbol{v}}$ is not defined since $\hat{\boldsymbol{v}}$ is not scalar. As alternatives, we can consider the expected RMSD if we desire a scalar summary of the whole estimator's behaviour or produce a vector with an MSE value for each pixel. We note that either of these measures will be susceptible to the column rotation described above. These measures will likely prefer an estimator that, for example, blurs the image over one that simply rotated the columns ahead by one. This seems to indicate that as a general-purpose evaluation of estimators in imaging, we do not want to use the 2-norm.

However, if we restrict the problem we find that metrics based on the 2-norm become more reasonable. A sufficient restriction is that we only attempt to evaluate independent

estimators. From our previous definition, we know that each voxel output from the estimator has a distribution that depends only on the true parameters of the voxel's distribution and has no relationship to its neighbours' values. This precludes shifts and similar 'perceptually-tricky' transformations because they require that an estimated voxel depend on neighbours' true values. Given this restriction, we can treat estimation as a separable problem, with each pixel being estimated independent of its neighbours. In this case, minimizing the MSE penalizes estimators that are occasionally very far off in favour of those that stay in a narrower band around the true value. In image terms, reducing MSE at each voxel decreases the probability that some of the voxels are very different from the true voxels.

3.3 Quantifying Quality with Task-Free Observer Experiments

A major complaint with using MSE as the measure of quality is that it is not justified based on how the viewer perceives the image. Even in the case where we have assumed each voxel's estimate is independent, we have no clear evidence from human vision that the 2-norm is the preferred error metric. We might, in fact, prefer images that have a very particular distribution of error. Furthermore, what sorts of images we prefer very likely depends on what we are trying to do with them and our experiences looking at images of a particular type. Thus, what a radiologist desires in an estimated family photograph might not be the same as what she or he wants in an estimated MRI slice image.

One solution to this problem is to ask radiologists to score the quality of images produced by different estimators. These experiments normally consist of a display with two images, one being constructed from the true image and the other being produced by one of the estimators under comparison. Radiologists are asked how similar the estimated image is to the true image. By comparing how the radiologists score a sample of images produced by the estimators, we can quantify the quality difference between the two estimators.

The principal difficulty with this approach is that, like the ideal Monte Carlo experiment outlined above, it requires a panel of radiologists. Although we have alleviated the need for ground truth and the production of radiological reports, this demand for radiologist time is still not practical in, for example, a research setting where many new image processing algorithms are being regularly developed and compared. To address this problem several models of the perceptual difference between two images have been proposed [21, 87]. The common idea behind most of these difference measures is to construct a function representing

the contrast sensitivity of the human visual system. These models attempt to determine whether two images can be distinguished by an average human 50% of the time. Such a difference is called a *just noticeable difference* (JND), and these can be in turn be used as units of difference.

Having quantified the difference between two images, these models can then be used to measure the difference between the output of an estimator and ground truth, thus providing a quantitative value of quality. The Case-PDM model (a perceptual difference model developed at Case Western Reserve University) has been used in several experiments studying MRI image quality. Among the many available perceptual difference methods, the Case-PDM has been shown to best match radiologists' quality scores in a variety of MRI pulse sequences [68, 54].

3.4 Quantifying Quality with Task-Based Observer Experiments

Perhaps the most common complaint leveled against the image quality measures described in sections 3.2 and 3.3 is that they fail to evaluate whether a modification measurably improves performance in the task of which the image is a part. Since successful completion of the task is the end goal of the image, it is reasonable to construct an experiment where only the estimator is varied and measure success in the task. The changes in task performance can then be used to quantify the quality difference between estimators. By far the most common use of MRI images is visual inspection by a radiologist. Through their training, radiologists develop specific techniques of observation and complex mental models of anatomy and disease to support their visual search and decision [39]. Since our images will be observed by the radiologist to produce the final outcome (a radiological report), we could measure the quality of the process's output, rather than measure the quality of the image that we provide as input. We can then judge the quality of an imaging system by the quality of the output a radiologist produces when using it.

Based on this idea, image quality in many medical modalities has been quantified based on their ability to support radiologists in a diagnostic task. To perform the experiment suggested above, we would ideally evaluate the correctness of the radiologists' reports produced when looking at the output of a given estimator. However, as we noted at the beginning of this chapter, in practice such an experiment is not feasible for most researchers. Instead, we

must usually accept a less realistic task in order to conduct our experiments. In general, all of the tasks considered for image quality measures are detection tasks. In these tasks, the observers must in some way either detect an object or express their confidence that they have detected an object in a given image. The choice of these tasks is related to a particular model of the cognitive process involved in detection and the influential receiver operator characteristic (ROC) analysis technique that results from this model.

3.4.1 ROC analysis

In the 1950s the American military began developing techniques to quantify the utility of various radar equipment using a task-based form of experimentation. The task performed by a radar operator is actually quite similar to that of a radiologist, involving a visual search of the display and then classification of each feature of interest into positive/enemy/disease or negative/friendly/healthy. The results of an experimental trial can be described by the fraction of true-positives (TPF) and false-positives (FPF) reported, be they enemy aircraft or tumors. This is known as ROC analysis. Due to these task similarities, the tools for radar quality have been adopted by the medical imaging community for their own studies [56].

In this scheme, a viewer is assumed to have some internal criteria by which she or he determines a diagnosis. We can model this process by describing these criteria as a function the viewer applies to each image. This function produces a test statistic, λ, on a continuous scale representing the viewer's diagnosis and their confidence in it. Thus, the middle of the scale represents the viewer being unsure of the state of the imaged patient while scores at the extreme ends of the scale represent confident determination of a positive or negative case.

The viewer is hypothesized to turn their rating into a final diagnosis by comparing the score to a threshold value, t. If the viewer makes their decisions more lax or more strict they will affect the number of positive diagnoses they produce on a set of images. Let us say a radiologist looks at a set of images with her or his internal threshold set to $t = t_1$ and produces some number of illness diagnoses. If the radiologist then looks at the same images but increases her or his threshold to $t = t_2$, $t_1 < t_2$, fewer patients would be diagnosed as ill since each image's score would be the same but fewer would be over the threshold.

Since we are interested in the TPF and FPF, we can describe the behaviour of t with two distributions: $p(\lambda; +)$ and $p(\lambda; -)$, the PDFs of the scores λ in the condition where the image

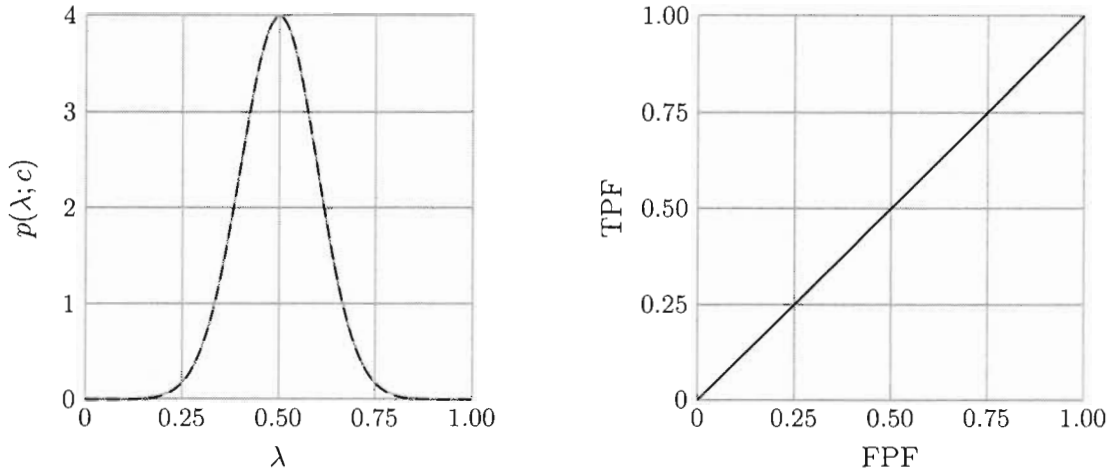

Figure 3.2: ROC model of a useless imaging system. *Left:* PDFs of the score function λ when in the positive/disease case $p(\lambda; +)$ and the negative/healthy case $p(\lambda; -)$ are plotted in solid and dashed lines respectively. Note that they overlap in this plot since the PDFs are identical. *Right:* ROC curve. Since the TPF and FPF distributions completely overlap, changing the threshold will affect the number of positive diagnosis, but at every threshold diagnosis still has a 50% chance of being correct.

is in the disease-positive or disease-negative state respectively. Given these distributions, we can calculate the probability of correct and incorrect positive diagnoses for a given threshold t as $P(\lambda > t; +)$ and $P(\lambda > t; -)$ respectively. The curve produced by plotting the true-positive probability on the y axis versus the false-positive probability on the x axis over the full range of possible t produces the ROC curve.

If the TPF and FPF are identical at all t, our test result is no better than chance. For this to be the case, we must have $p(\lambda; -) = p(\lambda; +)$ (left side of figure 3.2). Tracing out all the possible values of t, this produces a perfectly diagonal line as a ROC curve (right side figure 3.2). If the TPF and FPF distributions do not overlap at any threshold, then we have a test that diagnoses perfectly. In practice we expect the TPF and FPF to fall between these two cases, having different distributions that overlap somewhat (left side of figure 3.3). A ROC curve that is above the diagonal represents a test of this sort (right side of figure 3.3)

One of the primary difficulties in using ROC curves to compare different estimators is deciding which curve, and thus which estimator, is better. If one estimator's ROC is higher

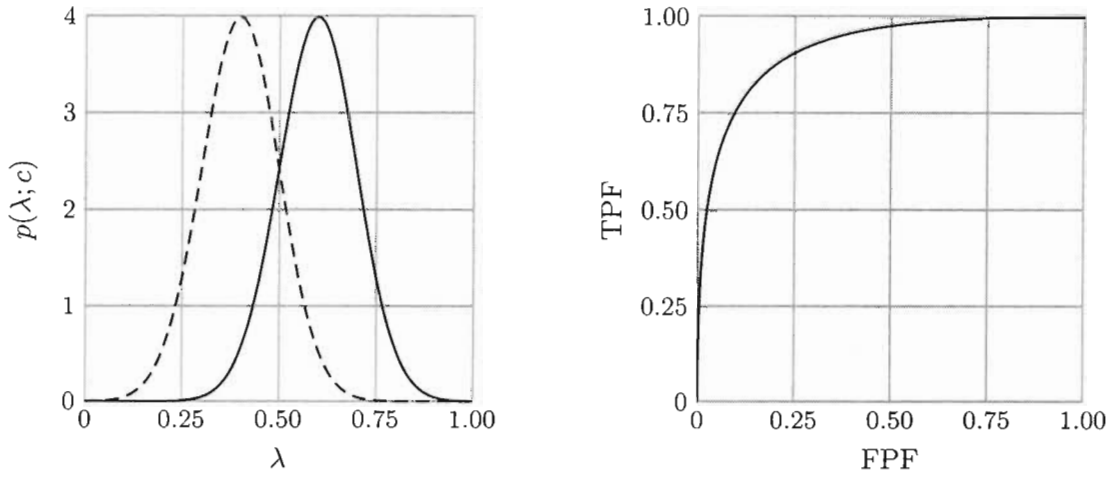

Figure 3.3: ROC model of a useful imaging system. *Left:* PDFs of the score function λ when in the positive/disease case $p(\lambda; +)$ and the negative/healthy case $p(\lambda; -)$ are plotted in solid and dashed lines respectively. The curve being above the diagonal results from there being more true-positive diagnoses than false-positive diagnoses across all possible threshold points.

than its competitor at every t, then the superiority is clear. However, in cases where the curves intersect another metric is needed. A variety of schemes have been developed for comparing ROC curves based on single-valued metrics [53]. In our work, we will focus on the area under the ROC curve (AUC) as our metric of quality. As we will see shortly, this metric has a particular appeal given the form of our experiments.

3.4.2 Choosing Tasks and Observers

The most direct experiment based on the above model of cognition is to provide observers with a series of images and ask them to select a score for the image out of some range to indicate how confident they are that the image is in the positive/disease category. Having collected this data we can then make post-hoc decisions on behalf of the observer by applying a chosen threshold. If we vary the threshold, we can trace out the ROC curve for the observer viewing the presented images [53].

Another task commonly used as a surrogate for full diagnosis is n-alternative forced-choice (n-AFC) signal known exactly (SKE) detection. In this task an observer, ideally a

radiologist, is simultaneously shown n images produced by the estimator or a simulation thereof. All of the images contain the anatomy, noise, and other normal aspects of the imaging system's output. However, one image also contains a target feature. To assist detection, the observer is shown where the feature would be in each of the n images if it was present. They are also, separately, shown the feature so that they know what they are looking for. The job of the observer is then to give their best guess which one of the n images contains the feature. A natural measure of quality resulting from this experiment is the percentage of correct guesses under each imaging system. In terms of the model of cognition outlined above, it is assumed that the observer gives each of the n-images a score and then selects the one with the greatest score as her or his choice [53].

The choice of the n-AFC SKE experiment avoids the problems associated with asking an observer to score images on a scale (e.g., crowding all their responses at one end of the scale). It is also conveniently the case that the percentage correct in the 2-AFC is an estimator for the AUC assuming the observer applies the score-and-threshold model laid out above [9]. Thus, using a 2-AFC experiment we can easily calculate the the AUC as a quality metric without asking the observers to score each image.

Just as in the case of task-free models, the requirement of a panel of radiologists to judge the quality of every imaging system under research is impractical in most research settings. However, a variety of model observers have been developed that can provide scores related to the confidence of having detected a signal, simulating the model of a human's internal process. If we are performing an n-AFC detection task, we can take the model's choice to have been the image with the higher score. This then allows us to apply the same analysis as for human choices.

A variety of models have been presented, many having had strong correlation with human results in initial experiments testing specific modalities (for a summary, see [56, 23, 1]). The models considered are generally linear, in the sense that the score assigned to an image is given by the dot product of a weight vector $\boldsymbol{w}$, based on the signal to be detected, and an input vector representing the image. However, the image and feature data are often transformed through a non-linear model of the perceptual system (*channelized*) before being used as input to the linear score function. The models are then taken to be linear in the channel responses.

The restriction to linearity is partially based on the strong optimality results available in the statistical decision theory. However, one recent experiment that tried to use non-linear

models inspired by the human visual system found little improvement over linear models for a 4-AFC detection task with compressed medical images [94]. Based on the availability of well-developed linear models and the lack of positive results for any proposed non-linear models, we will focus on linear models for task-based detection. In section 3.5 we will go into explicit detail about two of these models, the channelized Hotelling observer (CHO), and the non-prewhitening eye-matched (NPWE) observer, when we consider their application to MRI data.

3.4.3 Comparison of task-based and task-free observer quality measures

The suggestion of using a reduced task to simulate the full task of radiological diagnosis is, in some ways, a worrisome trade-off. In particular, criticizing task-free quality measures outlined in section 3.3 for being unrelated to task success may be somewhat disingenuous if they are to be replaced by measuring success on an unrelated task. In essence, since we cannot measure success on the full, realistic task, we are choosing between trusting a radiologist's judgement about which image they prefer, or asking the radiologist to perform a simpler task and suggesting the results generalize to the full diagnostic task.

The difficulty in choosing between these methods is particularly acute because it is difficult to structure an experiment that evaluates them equally. As yet, there are no established methods for converting a result from a tool like Case-PDM into a decision in a given n-AFC SKE instance or vice versa. However, one experiment has attempted to determine which method produces better results in the parameters of a lossy compression algorithm for medical images [26]. In this study the NPWE observer performing a 4-AFC task was compared with a task-free metric known as DCTune (based on the discrete cosine transform and a model of human perception) over a collection of images to find the compression settings that optimized performance under each model. After compressing images according to the recommendations of each measure, the results were shown to human observers who performed 4-AFC tasks. The results indicated that the settings recommended by the NPWE resulted in better detection by humans observing compressed images than when settings used by DCTune were recommended.

Generalizing from this result would require more experiments, but there is also an additional concern: the task being used by humans to measure post-compression quality is still a reduced task, and thus may not be indicative of performance in full radiological viewing.

However, at the least, this experiment suggests that task-specific models can be more accurate predictors of human performance in n-AFC task than a general perception model. This suggests that there likely are limits to a general perception model's ability to measure quality for every specific task that might be performed by a radiologist.

3.5 Our Task-Based Experiments of MRI Estimator Quality

There is a large body of literature using human observers and mathematical models of human response on simple detection tasks using x-ray [25, 24] and nuclear-medicine images [67]. However, there had not previously been any studies using task-based quality measures on MRI data before our work on this problem [80, 81]. Although there are many possible aspects of MRI that could be evaluated using task-based measures, we were particularly interested in whether phase-corrected real reconstructions would improve image quality. We felt that experiments with human observers could provide a more solid indication of the value of phase correction than the previous theoretic result [13]. In addition to comparing phase-correction with the current standard magnitude reconstruction, we also opted to evaluate the quality of images produced by a widely cited wavelet thresholding scheme for MRI images [59] (we will consider this estimator in more detail in chapter 5).

We proposed using both human and model observers to compare the three estimators of signal magnitude in weighted proton imaging data. We used high-SNR MR images as backgrounds combined with complex $\mathcal{N}(0, \sigma_Q)$ noise to produce our simulated low-SNR raw MRI measurements. Our human observers were volunteers without radiological experience and we compared their performance against a CHO [57, 1] and a NPWE observer [17] in a 2-AFC SKE detection task.

Using this configuration, we performed two separate experiments. The parameters of the second experiment were varied subtly from the first, partially based on lessons learned in running the first experiment, but also in order to explore how the estimators affected different types of images. In the following descriptions of the experiments we will refer to these as experiment one and experiment two when specifying parameters in order to keep the explanations clear.

In experiment one, we used only a CHO and set its parameters with knowledge of all the human results. While this allows the overall fidelity of the model to be considered, it is not a realistic setup for testing how a model observer might actually be used to evaluate

novel estimators. If we are required to perform human studies to tune the parameters of the model for every new experiment, then we could just as well use the results of the human study and skip the model. To explore this idea, in our second experiment we set the parameters of the model observers with information only about human observers' performance on magnitude MR images. We then tested the model observers' ability to extrapolate to novel estimators. Experiments structured this way would allow researchers developing new estimators to quantify the results of their work without needing new human studies for each new estimator.

3.5.1 Synthetic images

When attempting to locate a target feature in an MR image there are three major sources of distraction: non-target patient anatomy, artifacts, and thermal noise due to resistance. In terms of our signal model, equation (2.1), the anatomy, artifacts, and target feature combine to form the image signal, s, while the noise $q_r + \mathrm{i}q_i$ is additive and drawn from a complex $\mathcal{N}(0, \sigma_{\mathbf{Q}})$. By producing synthetic images with both of these components we propose that target feature detection in our synthetic images will have approximately the same results as similar tasks in clinical MR images when using the estimators under consideration.

Expanding from a single voxel m to a matrix $\mathbf{M}$ representing an image, we create a complex-valued image without a target feature via

$$\mathbf{M} = \mathbf{B} + \mathbf{Q_r} + \mathrm{i}\,\mathbf{Q_i}\,, \tag{3.7}$$

where $\mathbf{B}$ is a real-valued background anatomy image, and $\mathbf{Q_r}$ and $\mathbf{Q_i}$ are images whose pixels are each drawn from $\mathcal{N}(0, \sigma_{\mathbf{Q}})$. Similarly, we produce a complex-valued image with a target feature using

$$\mathbf{M} = \mathbf{T} + \mathbf{B} + \mathbf{Q_r} + \mathrm{i}\,\mathbf{Q_i}\,, \tag{3.8}$$

where $\mathbf{T}$ is the real-valued image containing only the target.

Equations (3.7) and (3.8) do not include signal phases. This is a substantial deviation from the model in equation (2.1), but can be be justified by considering the processing that will be applied to these images. The result of the magnitude transform on either the feature-present or feature-absent images defined above will be invariant to signal phase since the magnitude transform discards phase. Similarly, because the wavelet transform being considered operates on images after the magnitude transform, its result is also correct without needing a simulated signal phase.

The phase-corrected real reconstruction does rely on the signal phase ϕ either being known or estimated, and thus potentially on a simulated signal phase. However, we note that in the best case, the estimate of ϕ is exactly the same as the true ϕ and so we can simulate the best case result of phase correction with

$$\mathrm{Re}(\mathbf{M}) = \mathbf{B} + \mathbf{Q}_r \,, \tag{3.9}$$

in the target-feature-absent case and

$$\mathrm{Re}(\mathbf{M}) = \mathbf{T} + \mathbf{B} + \mathbf{Q}_r \,, \tag{3.10}$$

in the target-feature-present case. If this best-case estimate of phase-corrected real reconstruction cannot show an improvement in observer detection, then we should be doubtful of any realistic implementation providing improvement.

We also note that if a phase estimation scheme over-fits the phase measurement (i.e., if the phase due to noise is fit, instead of just the signal phase) the result will be an approximation to the magnitude transform. Alternatively, if the phase estimation under-fits the data there will be spatially varying signal intensity but the noise power will remain the same. This will have the same effect as lowering the target feature intensity relative to the thermal noise. Thus, while we do not simulate the failure conditions directly in our experiment, the effects of both these types of failures will be discernible from the experiment results because we cover both the magnitude transform and a range of target feature intensities and thermal noise powers.

Our target feature, $\mathbf{T}$, is an anti-aliased circular object located in the center of the feature-present image. If we index pixels with coordinate vectors $\boldsymbol{x}$, then this target feature image matrix is given by

$$\mathbf{T}[\boldsymbol{x}] = \begin{cases} b & \text{if } \|\boldsymbol{x} - \boldsymbol{z}\| \leq w \\ b(1 - \|\boldsymbol{x} - \boldsymbol{z}\| + w) & \text{if } w < \|\boldsymbol{x} - \boldsymbol{z}\| < 1 + w \\ 0 & \text{otherwise} \end{cases} \tag{3.11}$$

where b is the amplitude of the feature, $\boldsymbol{z}$ is the index of the image center, and w controls the width of the feature. For our experiments we set $w = 3$ which, given the viewing distance and image size discussed in section 3.5.2, is equivalent to an anatomical feature with a diameter of 6 mm. In experiment one, we selected $b \in \{\frac{1}{20}, \frac{1}{12}, \frac{5}{36}\} \simeq \{0.05, 0.083, 0.139\}$

for each image. In experiment two, we used $b \in \{\frac{3}{100}, \frac{1}{20}, \frac{1}{12}\} \simeq \{0.03, 0.05, 0.083\}$, slightly dimmer than in experiment one.

The complex-valued thermal noise was simulated by first selecting $\sigma_{\mathbf{Q}}$ and then randomly generating two 128×128 pixel noise images for each synthetic MR image. Each pixel in these noise images was sampled from $\mathcal{N}(0, \sigma_{\mathbf{Q}})$. One of these images was taken as $\mathbf{Q_r}$ and the other as $\mathbf{Q_i}$. In both experiment one and experiment two we selected $\sigma_{\mathbf{Q}}$ from $\{\frac{9}{200}, \frac{3}{40}\} \simeq \{0.045, 0.075\}$.

To simulate distracting anatomy, $\mathbf{B}$, we used regions of slices from high-SNR MR head images of healthy volunteers. These volunteers were scanned using a 3D inversion recovery pulse sequence on a Philips Gyroscan Intera 3.0 Tesla MRI scanner. Each volume was reconstructed using the magnitude transform to give real-valued images. These real-valued volumes were then sliced along the axial, coronal, and sagittal directions to produce a library of 2D images. Each slice was then cropped into separate 128×128 pixel images which formed a library of candidate backgrounds for our experiments.

In experiment one, we selected images that had anatomy in the central part of the image in order to ensure the signal would be overlaid on the anatomy. Selections were made by summing the portion of each candidate background that would overlap with the signal. If the sum was above a chosen threshold, we assumed the majority of the target would overlap with anatomy. Images above the threshold were retained as backgrounds for experiment one and images below the threshold were excluded. The threshold was chosen by trial and error to produce a set of backgrounds for experiment one with the desired properties.

In experiment two we selected images that effectively had no signal where the target feature would be placed, but did contain bright anatomy immediately next to the target. We did this by summing the region that the target would overlap and comparing it with a threshold sufficient to exclude more than one or two pixels of anatomy intersecting the regions. Thus, our target was almost always placed on top of air, bone or some other region that emits little or no signal. We then summed the region immediately outside the target overlap and ensured it was above a threshold that guaranteed the presence of anatomy. As in experiment one, the two thresholds were chosen by trial and error in order to get a large set of backgrounds that met our description. The decision to locate the target next to anatomy in experiment two was made so that targets would only directly contrast against the dark background, but there would still be distracting anatomy in the observer's visual field.

In both experiments, backgrounds with the desired configuration of anatomy were normalized so that their pixel intensities were on the range $(0, 1)$. As noted in section 2.1, the background images produced by cropping and normalization had Rician noise because the magnitude estimator was used to produce the greyscale values. Since we used these images as our real-valued data, this Rician noise will have been added to the normally distributed $\mathbf{Q_r}$ and included in all of the synthetic image estimates. However, because our anatomical images were scanned at 3 Tesla, the magnitude images, after being normalized to the range $(0, 1)$, had a standard deviation of approximately 0.002 measured in regions of air. Since the lowest standard deviation, $\sigma_{\mathbf{Q}}$ used for our simulated thermal noise is more than 20 times greater than the inherent noise of our anatomical backgrounds, it is unlikely that the spurious noise included in the anatomical images had any effect on our results.

For experiment one, 672 images were produced in each of the 36 possible conditions (three estimators, three signal powers, two noise powers, target feature present or absent). In figure 3.4 we show one feature-present anatomical background in all 18 possible conditions of experiment one. For experiment two, 560 images were produced in each of the 36 conditions. In figure 3.5 we show one feature-present anatomical background in all conditions for experiment two.

The choice of parameters in both experiments was made by using pilot studies to ensure that the human responses would cover the range of responses from 50% correct (the task is so hard the observer is simply guessing) to 100% correct (the task is so easy the observer is always right). However, it is then important for us to verify that these parameters produce realistic images. To quantify our synthetic images' quality using the same calculations commonly used to describe clinical MRI, we define the peak signal-to-noise ratio (PSNR) as the ratio of the peak signal to the Rician noise standard deviation in a region of air [32],

$$\text{PSNR} = \frac{\max_{\boldsymbol{x}} \mathbf{B}[\boldsymbol{x}]}{\sigma_{\mathbf{Q}}\sqrt{(2 - \pi/2)}}. \tag{3.12}$$

Due to our normalizing all the anatomical backgrounds to the range (0,1), we know that $\max_{\boldsymbol{x}} \mathbf{B}[\boldsymbol{x}] = 1$. Thus, for images where $\sigma_{\mathbf{Q}} = \frac{9}{200}$ we calculate PSNR $\simeq$ 33.92 and for $\sigma_{\mathbf{Q}} = \frac{3}{40}$ we have PSNR $\simeq$ 20.35 These are a little low, but not implausible PSNR values for clinical MRI.

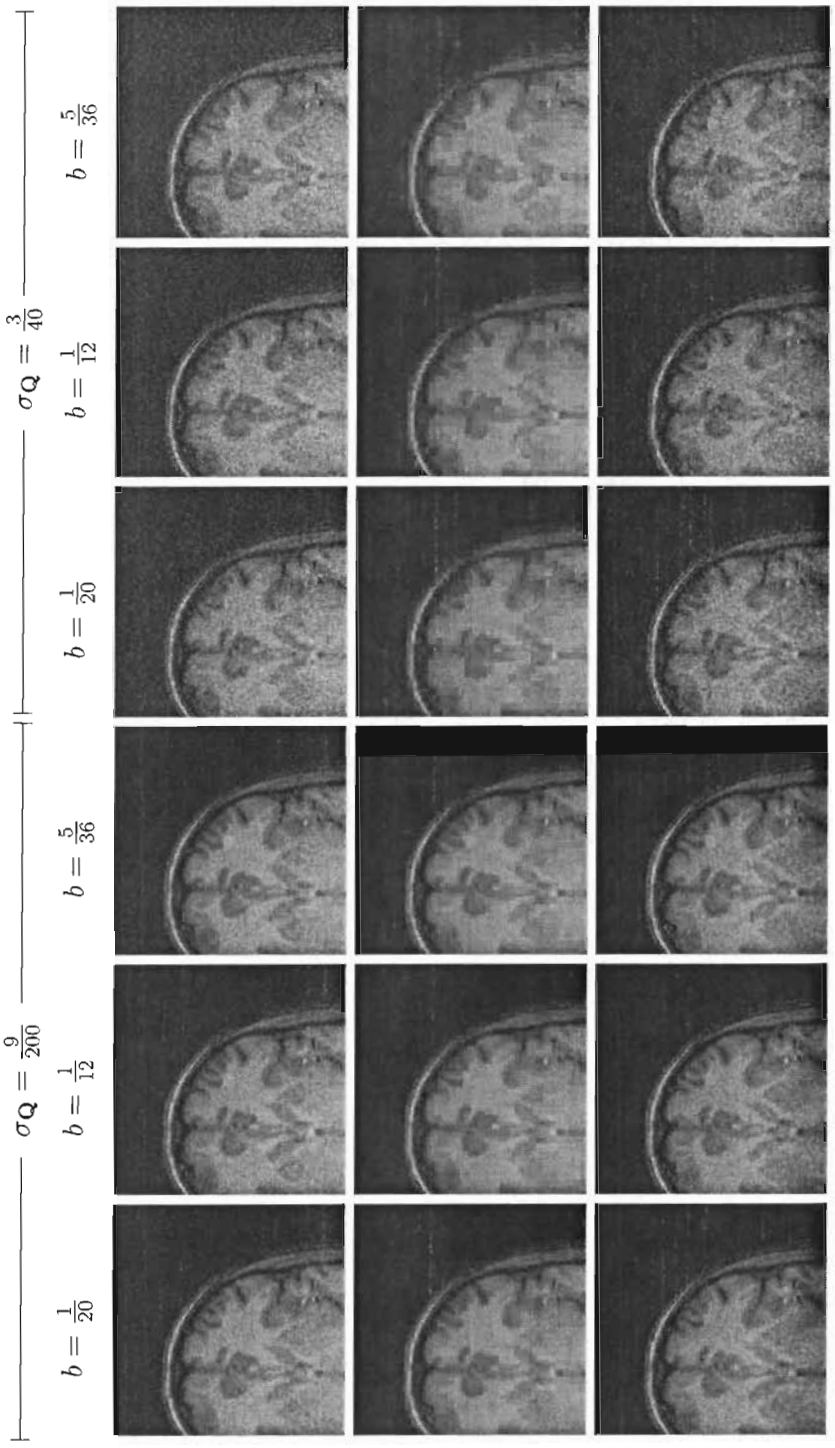

Figure 3.4: Examples of feature-present synthetic images for all of the possible 18 experimental conditions in experiment one. The first row was produced with the magnitude estimator, the second row with the wavelet estimator, and the third row with phase-corrected real reconstruction. Each column was produced with a different simulated signal intensity, as specified at the top. The first three columns were produced using $\sigma_{\mathbf{Q}} = \frac{9}{200}$ and the last three with $\sigma_{\mathbf{Q}} = \frac{3}{40}$. The target feature (a small bright circle) is most visible in column three, near the top right surface of the patient's brain, located in the exact center of the image.

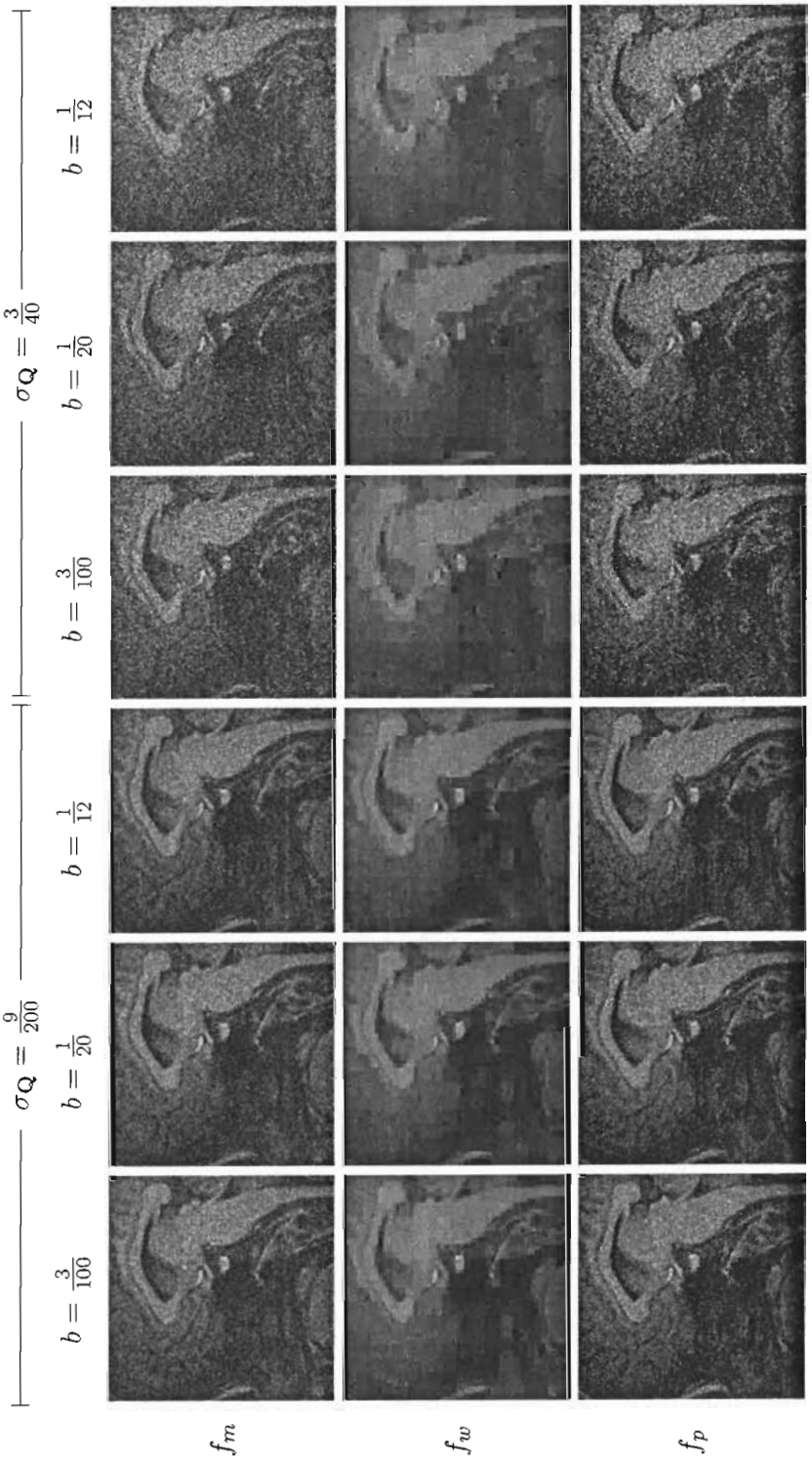

Figure 3.5: Examples of feature-present synthetic images for all of the possible 18 experimental conditions in experiment two. The first row was produced with the magnitude estimator, the second row with the wavelet estimator, and the third row with phase-corrected real reconstruction. Each column was produced with a different simulated signal intensity, as specified at the top. The first three columns were produced using $\sigma_{\mathbf{Q}} = \frac{9}{200}$ and the last three with $\sigma_{\mathbf{Q}} = \frac{3}{40}$.

3.5.2 Human observers

Sixteen volunteer observers without any previous radiological training were recruited to participate in each experiment, with no overlap between subject groups. We suggest that the use of subjects without radiological training was acceptable for this experiment for two reasons. First, the task itself, while similar to a radiologist's task, does not require any prior knowledge of anatomy due to the highly synthetic nature of the embedded target. Second, it has been suggested that radiologists are not better than average image viewers at the basic task of perception, but are more successful because they bring substantial mental models to bear in refining their search of the image and interpreting its contents [39]. Since our task focused almost entirely on how perceptible the targets were, requiring no search or medical interpretation, the use of non-radiologist subjects should not be a significant detriment to this type of experiment.

The experimental software presented participants with three images aligned horizontally (see figure 3.6). The center image showed the target feature, and the two exterior images represented choices in the 2AFC test. Since this was an SKE task, crosshairs were superimposed over the images in order to reduce the possibility of confusion about the target feature location. The crosshairs could be toggled on and off by the user to reduce visual distraction. Participants were instructed that in every display, one of the exterior images would contain the target feature and that they should use the mouse to click on whichever exterior image they felt most probably contained the target. They were allowed to take as long as they wanted to reach a decision on each image pair. Once a participant clicked on an exterior image, the screen was made completely black for 0.5 seconds, the mouse pointer was warped to the center of the screen, and then the next set of images was shown and the process repeated.

Each participant was given two training sets, composed equally of all 18 possible combinations of noise power, target feature power, and estimator. If the training took less than ten minutes, they were then instructed to wait until ten minutes had elapsed in order to ensure a constant dark adaptation time across all participants. After the training and delay, the participants then proceeded through 16 experiment sets composed equally of the 18 possible combinations for a total of 288 image pairs. In order to minimize order effects, the ordering of the image pairs was selected randomly for each participant from a constant distribution of all possible orderings. As the experiment lasted up to an hour, fatigue was

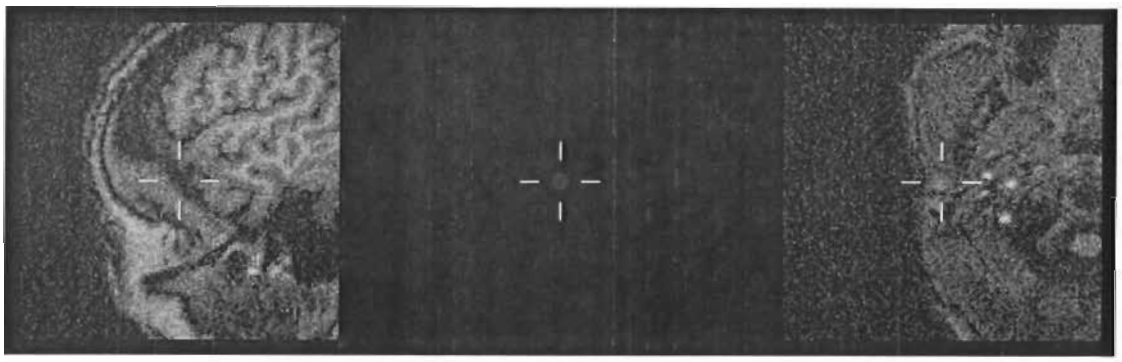

Figure 3.6: Example of the user interface used in the human observer experiments. In this case the target feature is in the right image and located on top of anatomy, as would occur in the images we used for experiment one.

reduced by displaying a black screen after every 18 image pairs and instructing users to take as long a break as they desired while the display was dark.

The experiments were conducted in a completely darkened room using a CRT monitor (SGI CMNB024B) as a display. The monitor's spectrum was measured in both the left and right image centers using a telephotometer (Photo Research PR-650) over a range of digital pixel values. The luminance was computed for each digital value using the CIE 1931 observer [92] (see figure 3.7). The luminance was not corrected to the DICOM standard, but we were satisfied that the critical areas at the center of the two images had sufficiently close luminance curves that our display could be considered sufficiently homogenous across its extent. All participants viewed the monitor from approximately 50 cm away while wearing any corrective lenses they would normally use for computer viewing. When displayed on the screen, the images each had a diameter of approximately 8.5 cm and so each occupied an angle of approximately 10 degrees from the viewer's eye position.

In order to display the images without distorting their brightness, we had to ensure that the pixels' values in each experiment's image set fell in the digital greyscale range (0,255). To achieve this, we first truncated all negative image pixel intensities at 0. The maximum pixel intensity in each experiment's image set was found: ~ 1.36 for experiment one and ~ 1.2925 for experiment two. All pixels' intensities were then scaled by the same value in each experiment set (~ 186.94 in experiment one and ~ 197.29 in experiment two) to ensure complete coverage of the digital greyscale range. By imposing a consistent scaling on all images, some (e.g., those with a low-intensity target feature and low noise power) did not

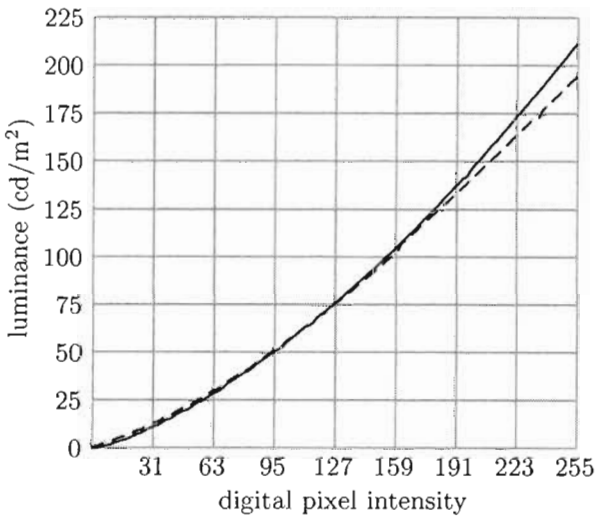

Figure 3.7: Luminance in cd/m^2 of the monitor (y-axis) plotted against the digital greyscale intensity sent to the video card (x-axis). The solid line is the luminance measured at the location of the left image's center and the dashed line is the luminance measured at the right image's center.

use the entire range of display intensities.

3.5.3 Model Observers

Channelized Hotelling observer

We applied the CHO [57, 1] with Gabor channels to our synthetic images. The CHO is the optimal linear two-class discriminator when the covariance of each class's channelized form is known. Gabor channels in particular were used because it has been suggested they are a useful approximation for the grating response of the human visual system [25, 24, 61]. The channels are defined by the response equation

$$\mathbf{G}[\boldsymbol{x}] = \exp\left(-4(\ln 2)f^2\frac{(\boldsymbol{x}-\boldsymbol{x_0})^t(\boldsymbol{x}-\boldsymbol{x_0})}{w_s^2}\right)\cos\left(2\pi f\{\cos\phi,\sin\phi\}(\boldsymbol{x}-\boldsymbol{x_0})+\beta\right). \quad (3.13)$$

where $\boldsymbol{x_0}$ is the image center, f is the central frequency of the filter in cycles per pixel, w_s is the filter width in octaves of f, ϕ is the angle of the filter, and $\beta \in \{0, \frac{\pi}{2}\}$ determines if it is odd or even. We have used a setup with forty channels, based on the example of Eckstein [25, 24], with $w_s = 0.8825$ and $\phi \in \{0, \frac{2\pi}{5}, \frac{4\pi}{5}, \frac{6\pi}{5}, \frac{8\pi}{5}\}$. To compute the central frequencies, note that each pixel subtends $\frac{5}{64}$ degrees from the viewer's eye position. We would like our filters to have frequencies of 2, 4, 8, or 16 cycles per degree. Converting this to cycles per pixel gives $f \in \{\frac{5}{32}, \frac{5}{16}, \frac{5}{8}, \frac{5}{4}\}$. The $16,384 \times 40$ channel matrix, $\mathbf{C}$, is produced by rearranging each channel, $\mathbf{G}$, as a $16,384 \times 1$ vector and making them each a column of $\mathbf{C}$. We can then compute the 40×1 channel response vector, $\boldsymbol{u}$ of image $\mathbf{M}'$ by rearranging the image into the $16,384 \times 1$ vector, $\boldsymbol{y}'$ and setting

$$\boldsymbol{u} = \mathbf{C}^t\boldsymbol{y}', \quad (3.14)$$

where $\mathbf{C}^t$ is the transpose of $\mathbf{C}$.

In order to derive the CHO for each of the 18 experimental conditions in each experiment, we must compute the specific covariance matrix, $\mathbf{K}_{\boldsymbol{u}}^c$, for each condition, c. We first note that

$$\mathbf{K}_{\boldsymbol{u}}^c = \frac{1}{2}(\mathbf{K}_{\boldsymbol{u},-}^c + \mathbf{K}_{\boldsymbol{u},+}^c) + \mathbf{K}_{\boldsymbol{\epsilon}}^c \quad (3.15)$$

where $\mathbf{K}_{\boldsymbol{u},-}^c$ and $\mathbf{K}_{\boldsymbol{u},+}^c$ are the covariance matrices of the channel responses in condition c's target feature-absent and -present cases respectively, and $\mathbf{K}_{\boldsymbol{\epsilon}}^c$ is the covariance matrix of the observer's internal noise process in condition c. This process is assumed to add noise independently to each response channel by sampling from a normal distribution with zero

mean and variance depending on the channel. Following the example of Eckstein, we define $\mathbf{K}_{\epsilon}^{c} = \alpha_{\text{CHO}} \,\text{Diag}\left(\frac{1}{2}(\mathbf{K}_{\boldsymbol{u},-}^{c} + \mathbf{K}_{\boldsymbol{u},+}^{c})\right)$ where Diag() zeroes all the off-diagonal elements of its argument and α_{CHO} is a proportionality constant that can be varied to reduce the absolute performance of the model observer [24].

This leaves the problem of determining $\mathbf{K}_{\boldsymbol{u},-}^{c}$ and $\mathbf{K}_{\boldsymbol{u},+}^{c}$. Given that we have closed forms for neither the pixel covariance of the anatomical backgrounds, nor the effects of the wavelet filter, we opted to estimate $\mathbf{K}_{\boldsymbol{u},-}^{c}$ and $\mathbf{K}_{\boldsymbol{u},+}^{c}$ from the synthetic data. Noting that $\mathbf{K}_{\boldsymbol{u},-}^{c}$ and $\mathbf{K}_{\boldsymbol{u},+}^{c}$ are each 40×40 matrices, we used the channel responses from 400 signal-present and signal-absent images respectively to estimate each for each condition. In experiment one we had 560 target feature-present and 560 target feature-absent images in each of the 18 experimental conditions, this left 160 image pairs for testing each condition. In experiment two we had 672 target feature-present and 672 target feature-absent images in each of the 18 experimental conditions, leaving 272 image pairs for testing each condition.

With the 18 $\mathbf{K}_{\boldsymbol{u}}^{c}$ estimates computed, we can determine the optimal channel weights, $\boldsymbol{v}^{c}$, for each condition, c, according to the Hotelling strategy as

$$\boldsymbol{v}^{c} = (\mathbf{K}_{\boldsymbol{u}}^{c})^{-1} \left(\langle \boldsymbol{u}_{1}^{c} \rangle - \langle \boldsymbol{u}_{0}^{c} \rangle\right) , \tag{3.16}$$

where $\langle \boldsymbol{u}_{1}^{c} \rangle$ and $\langle \boldsymbol{u}_{0}^{c} \rangle$ are the sample mean target feature-present and target feature-absent channel response vectors for condition c. From this we can write the template, $\boldsymbol{w}^{c}$, applied by the observer in condition c as

$$\boldsymbol{w}^{c} = \mathbf{C}\boldsymbol{v}^{c} . \tag{3.17}$$

This template can be used to calculate the response, λ, to a reconstructed image $\mathbf{M}'$. Reordering the 128×128 matrix $\mathbf{M}'$ to the $16,384 \times 1$ vector $\boldsymbol{y}'$ we write

$$\lambda = (\boldsymbol{w}^{c})^{t}\boldsymbol{y}' , \tag{3.18}$$

where $\boldsymbol{y}'$ was produced with conditions c.

If there is no internal decision noise (i.e., if $\alpha_{\text{CHO}} = 0$) then deciding the CHO's choice in a 2AFC experiment requires calculating the λ for each of the two image choices and then selecting the image with the larger score. However, when $\alpha_{\text{CHO}} \neq 0$ we must add the internal noise of the observer. Rather than compute a noise for each channel, we note that the effect of the channel decision noises is combined in the final response score. Thus, we can modify the computed response by adding a single sample, ϵ, from $\mathcal{N}(0, \sigma_{\epsilon})$ where

$$\sigma_{\epsilon}^{2} = (\boldsymbol{v}^{c})^{t}\mathbf{K}_{\boldsymbol{\epsilon}}\boldsymbol{v}^{c} . \tag{3.19}$$

Adding this noise sample gives our final estimate of the score a human observer would assign to the image:

$$\lambda' = \lambda + \epsilon \,. \tag{3.20}$$

As discussed above, the image with the greater λ' in each pair is considered the CHO selection in the 2AFC trial.

Non-prewhitening eye-matched observer

Non-prewhitening observers are named based on their inability to take advantage of the covariance matrix to decorrelate the noise and background present in the image. The choice to ignore the covariance matrix is made because it has been suggested that humans are poor at decorrelating noise when searching for a feature. While the basic non-prewhitening observer simply multiplies the known signal with the observed image to determine λ, the eye-matched version, like the CHO, attempts to include information about the human visual system. In particular, the NPWE filters the observed image to simulate the spatial contrast response of the human eye [17]. In our experiments, we construct our template (which is the same across all conditions) with the equation

$$\mathcal{F}(\mathbf{W})[\boldsymbol{u}] = \mathcal{F}(\mathbf{T})[\boldsymbol{u}] \left\| E\left(\frac{5}{64}\boldsymbol{u}\right)\right\|^2 , \tag{3.21}$$

where $\mathcal{F}$ indicates the discrete Fourier transform, $\boldsymbol{u}$ is a 2-dimensional location in frequency space, $\mathbf{W}$ is the matrix form of the template that will be reordered to produce $\boldsymbol{w}$, matrices subscripted by vectors represent the matrix element indexed by the vector, and

$$E(\boldsymbol{x}) = \|\boldsymbol{x}\|^{1.3} \exp\left(-\frac{13}{320}\|\boldsymbol{x}\|^2\right) . \tag{3.22}$$

is the sensitivity model across angular visual field suggested as an approximation to the human eye response. The fraction $\frac{5}{64}$ appears in equation (3.21) to account for the change of units from pixels to angle subtended in the visual field, based on our calculations in the previous section. We note that we do not need to compute a different template for each condition since changes in noise do not affect the template and changes in target amplitude produce only a constant factor change in every pixel of the template.

In the absence of internal noise, we calculate the NPWE's choice between two images exactly as with the CHO, by taking

$$\lambda = \boldsymbol{w}^t \boldsymbol{y}' \,, \tag{3.23}$$

and guessing the image with the higher λ contains the target. To add decision noise to the NPWE, we select α_{NPWE} and then use

$$\lambda' = \lambda + \epsilon\,, \tag{3.24}$$

where ϵ is a sample from $\mathcal{N}(0, \sqrt{\alpha_{\mathrm{NPWE}}})$.

3.5.4 Human Experiment Results and Discussion

We computed the percentage correct, Pc, for each participant in each of the 18 experimental conditions. We computed the median Pc over all the participants as well as the first and third quartiles. As we noted above, Pc is an estimator of the AUC from the ROC analysis model [9]. Taking advantage of this, we have plotted the first, second, and third quartiles of the estimated AUC based on human performance in experiment one (figure 3.8) and experiment two (figure 3.9).

The width of the first and third quartiles indicates substantial inter-subject variability. The resolution of our AUC quartile measurements is only $\frac{1}{16}$ because we opted to cover a variety of experimental cases and thus show each participant each condition only 16 times. Although this structure allows us to study patterns over a large number of conditions, it also means that we are unable to demonstrate statistically significant differences between the estimators given the small effect.

Experiment one

Our human observers did not show a measurable increase in performance when using the phase-corrected real reconstruction. The lack of effect here may be due the to fact that the Rician distribution of a magnitude image becomes very similar to that of a phase-corrected real image when $\frac{\mathbf{S}[x]}{\sigma_{\mathbf{Q}}} > 3$. Since most of our target features were added completely on top of bright anatomy in this experiment, these two estimators should be effectively the same in many of the tested cases.

Similarly, we do not show an improvement in signal detection using the wavelet transform, despite the improved RMSD between the wavelet estimated image and the true image when compared to the RMSD between magnitude and true images [59]. It is not clear if a specific task would be better suited to this estimator, since it has been described as useful generally for MRI. Clearly, the wavelet basis underlying the filtering algorithm could be

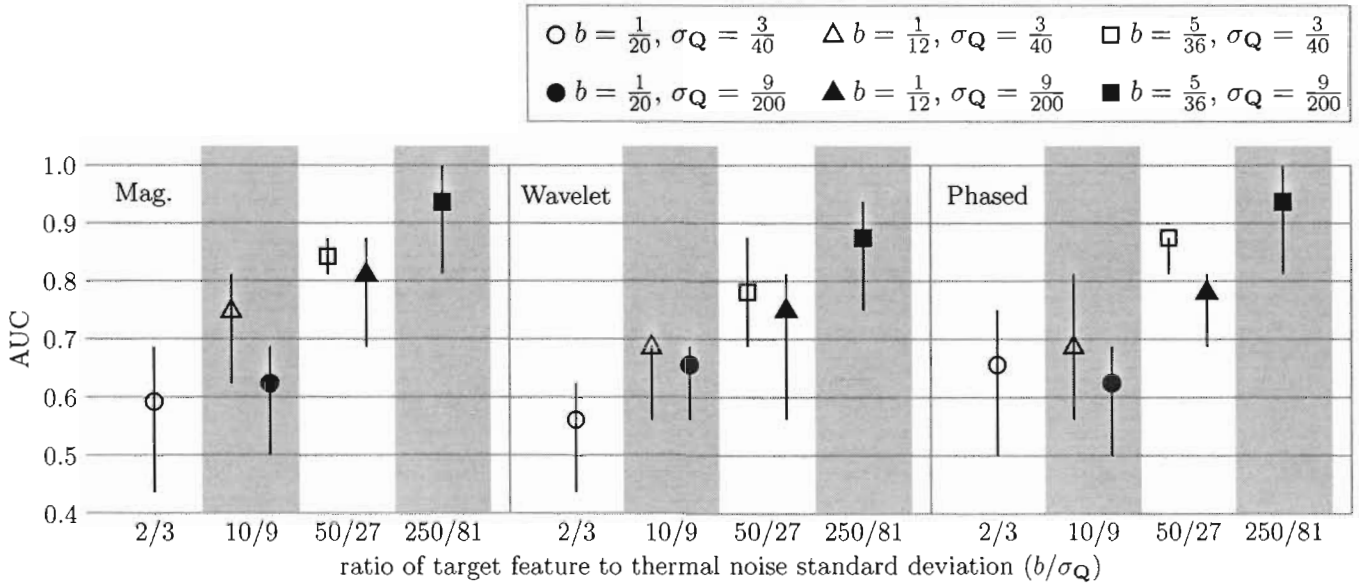

Figure 3.8: **Experiment One**: Plot of the median, first quartile, and third quartile of AUC for the human observers in each of the 18 experimental conditions. The y-axis is the AUC score. Each experimental condition is represented by a symbol (one of three shapes, either filled or unfilled) that locates the median AUC and two vertical lines that represent the first and third quartiles of the AUC. The symbol's shape represents a different configuration of target feature and noise power, as explained in the legend at the top-right of the chart. The x-axis is divided into thirds, with each third containing results from one of the estimation techniques as labeled in the top-left corner of each third. Within each third, the x-axis is further divided by the ratio of target feature to thermal noise standard deviation (complex feature-SNR). In some cases, two experimental conditions have the same feature to noise ratio (e.g,. $b = \frac{1}{12}, \sigma_Q = \frac{3}{40}$ and $b = \frac{1}{20}, \sigma_Q = \frac{9}{200}$) and so appear in the same band on the diagram.

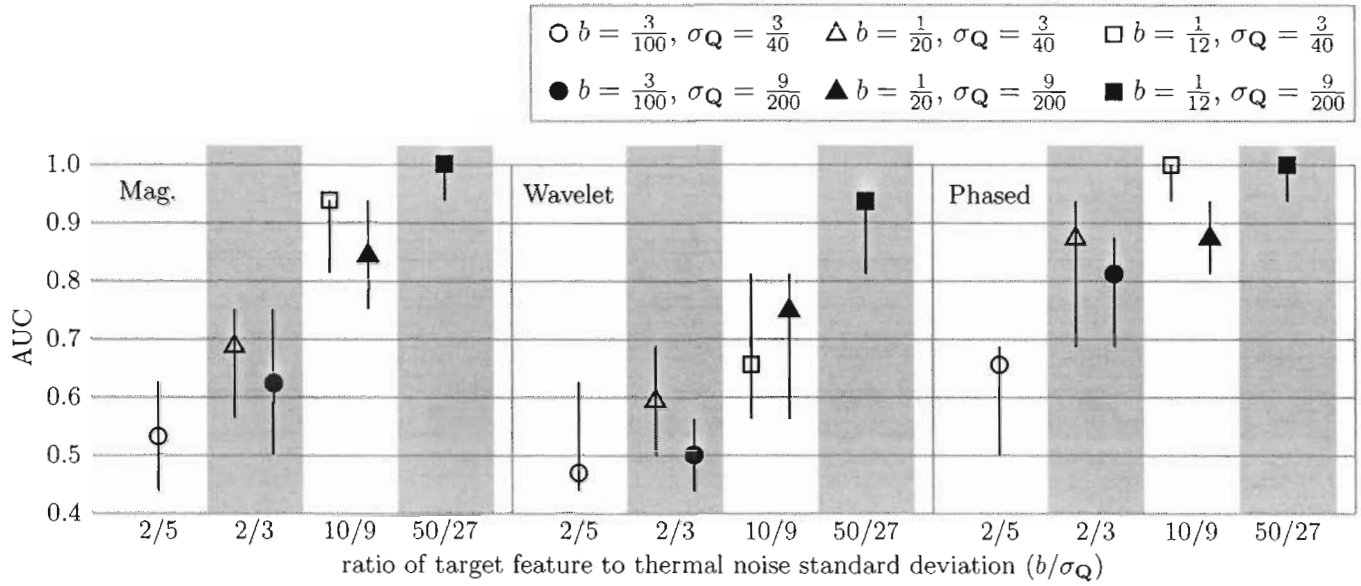

Figure 3.9: **Experiment Two**: Plot of the median, first quartile, and third quartile of AUC for the human observers in each of the 18 experimental conditions. The y-axis is the AUC score. Each experimental condition is represented by a symbol (one of three shapes, either filled or unfilled) that locates the median AUC and two vertical lines that represent the first and third quartiles of the AUC. The symbol's shape represents a different configuration of target feature and noise power, as explained in the legend at the top-right of the chart. The x-axis is divided into thirds, with each third containing results from one of the estimation techniques as labeled in the top-left corner of each third. Within each third, the x-axis is further divided by the ratio of target feature to thermal noise standard deviation (complex feature-SNR). In some cases, two experimental conditions have the same feature to noise ratio (e.g,. $b = \frac{1}{12}, \sigma_Q = \frac{3}{40}$ and $b = \frac{1}{20}, \sigma_Q = \frac{9}{200}$) and so appear in the same band on the diagram.

varied, and other wavelet processing algorithms could be implemented as well. However, our results reenforce the notion that claiming improved RMSD alone does not necessarily imply improved signal detection in MR images when evaluating regularized estimators.

Experiment two

Having failed to find any difference between phase-corrected real and magnitude images in experiment one, we moved the features to dark regions of the image in experiment two. This setup should provide the maximum difference we could expect to see between these two estimators. Although we have not shown statistically significant differences, the human results now appear to support the assertion that phase-corrected images improve detection compared to magnitude images when the signal is in the dim part of the image. This can be seen in figure 3.9 by comparing both the AUC of each signal/noise pair under the phase-corrected and magnitude reconstructions. In every case the phase-corrected reconstruction results in higher median AUC and in many cases the first and third quartiles have moved up substantially as well.

However, this experiment's results still conflict with the RMSD results that predicted the wavelet algorithm would enhance signal detection. However, we do note that the wavelet filter performed better at equivalent feature/noise/background ratios in experiment two than in experiment one. We suspect that the filter is more sensitive to edges in this part of the image and thus less likely to smooth them out.

Both experiments

Considering the magnitude and phase-corrected real AUCs, we note that in every case except one where two conditions share the same ratio of target feature amplitude to thermal noise standard deviation (vertical bands in figure 3.8 and 3.9), the condition with higher target feature intensity outperforms the condition with the lower target feature intensity. Since both the feature-to-noise ratio and anatomical background intensity were held constant in these cases, the only changes are the increase in target feature and noise intensity relative to the anatomical background. We hypothesize that this effect is due to the anatomical background obscuring the target feature more often when the feature is less intense. Combined with anecdotal comments from our subjects, this encourages us that using anatomical backgrounds to provide realistic distractors is important in studying feature detection.

Although the wavelet estimator produced similar results, it is unclear if this also demonstrates the effects of the anatomical background. The wavelet filter uses $\sigma_{\mathbf{Q}}$ as an input to control the amount of smoothing performed and so there is clearly a non-linear relationship, at least in theory, between AUC, the target feature intensity, and the simulated noise power. We cannot differentiate with our experiment whether the dominant effect in these conditions is the thermal noise or the smoothing artifacts.

3.5.5 Model Results and Discussion

We calculated the AUC for the CHO and NPWE by applying the template to each synthetic image, computing λ' using equations (3.20) and (3.24), and then totaling the number of correct selections to compute the Pc. We performed this operation for 50 separate instances of each model observer, and then computed the result of the mean observer in each of the experimental conditions. As in the human observer case, the Pc values were taken as estimates of the AUC.

Comparing models and humans – experiment one

In experiment one only the CHO was used. We were interested in testing how well a model observer could fit the human results across a variety of MRI estimators. The mean AUC of the CHO observers for each condition was fit to the mean human observer data by searching for the α_{CHO} that minimized the RMSD between the model and human results taking all 18 experimental conditions as a vector. We determined the optimal setting to be $\alpha = 3.2$ (RMSD $\simeq 0.00189$) by performing an exhaustive search on an initially coarse range of α values and then gradually refining the range. This computation took approximately three hours on a 1 GHz PowerPC G4. The mean AUC of the 50 model observers is shown in comparison to the mean human results in figure 3.10 (note that figure 3.8 displays the median human results while figure 3.10 displays the mean).

Overall, the CHO shows a very good match with the human study. There is no appreciable difference between the three estimators according to the model observer. We note that there was a small disagreement between the model and mean human observer on the ordering of some of the weaker signals. However, considering figure 3.8, we note that the inter-subject variability in these conditions was high as well. Additionally, there is a tendency for the CHO to slightly over-estimate mean human performance at the highest target

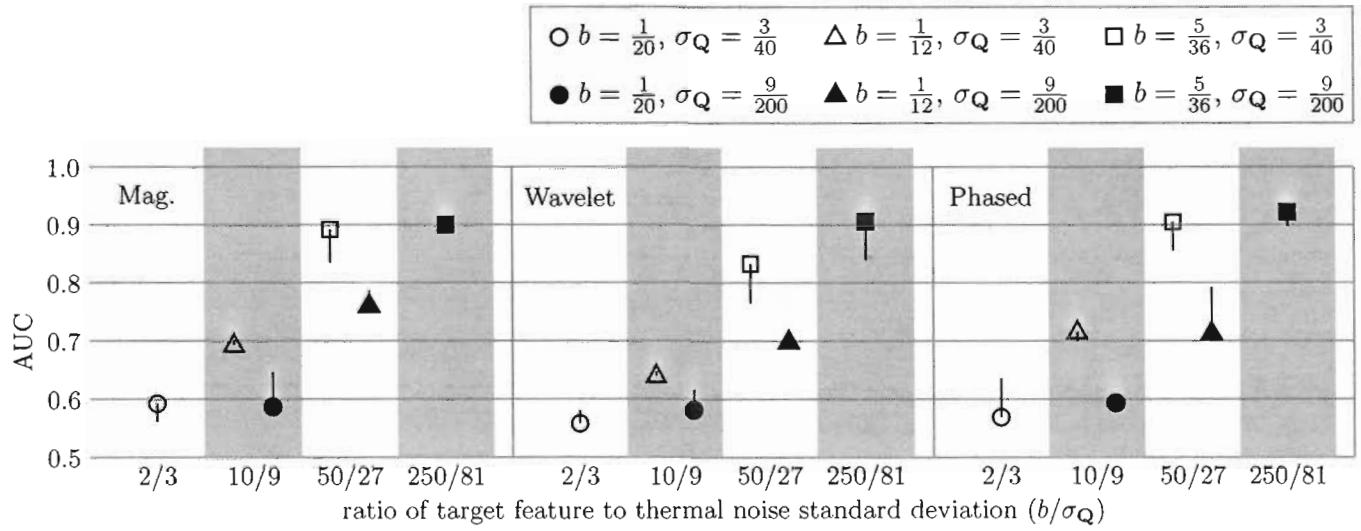

Figure 3.10: **Experiment One**: Plot of the mean AUC of the the model and mean AUC of the 16 human observers in each of the 18 experimental conditions. The y-axis is the AUC score. Each experimental condition is represented by a symbol (one of three shapes, either filled or unfilled) that locates the mean model observer AUC and one vertical line that represents the AUC of the mean human observer. Longer vertical lines indicate a greater mismatch between the model and human observers. The symbol's shape represents a different configuration of target feature and noise power, as explained in the legend at the top-right of the chart. The x-axis is divided into thirds, with each third containing results from one of the estimation techniques as labeled in the top-left corner of each third. Within each third, the x-axis is then further divided by the ratio of target feature to thermal noise standard deviation (complex feature-SNR). In some cases, two experimental conditions have the same PSNR (e.g,. $b = \frac{1}{12}, \sigma_Q = \frac{3}{40}$ and $b = \frac{1}{20}, \sigma_Q = \frac{9}{200}$) and so appear in the same band on the diagram.

feature intensity.

Using models to extrapolate human performance – experiment two

The mean AUC of the model observers was fit to the mean human observer data by searching for the α_{CHO} and α_{NPWE} that minimized the RMSD between the human and model results in the magnitude estimated experimental conditions. This was chosen to test the ability of the model observers to extrapolate to other estimators with different statistical properties given only human results in magnitude images. Using humans to calibrate the models based on only magnitude images was a feasible model for how an experimenter might proceed to test a new MRI estimator. With this approach the experimenter must only record human results once for the magnitude estimates of a given data set. Having determined this baseline model configuration, experiments could then evaluate as many other estimators as they desired using the calibrated models, as long as the statistical properties of the data set were unchanged.

We determined the best fit of the CHO to the mean human results on magnitude images to be $\alpha_{\text{CHO}} = 1.4$ (RMSD $= 5.19 \times 10^{-4}$ between human and CHO in magnitude images) using the same search process as in experiment one. This computation took approximately one hour on a 1 GHz PowerPC G4. Using the same technique, the optimal setting for the NPWE was found to be $\alpha_{\text{NPWE}} = 0.016$ (RMSD $= 2.92 \times 10^{-3}$ between human and NPWE in magnitude images). The mean AUC of the CHO and NPWE model observers is shown in comparison to the mean human results in figure 3.11 (again, note that figure 3.9 displays the median human results while figure 3.11 displays the mean). In both the phase-corrected and magnitude estimates, the model observers show a good match to the human study. In the wavelet estimated images, we find that the CHO overestimates human performance while the NPWE does not err consistently above or below. Noting that many of our human volunteers complained of "blurriness" in the wavelet estimated images, we suspect that the CHO may be more successful than humans when compensating for the smoothing due to its use of the channel covariance matrix.

3.6 Choosing an Evaluation Method for MRI Estimators

Having seen several approaches to evaluating the quality of an image produced by estimation, we must now decide which of these methods we will use for continuing our research. In

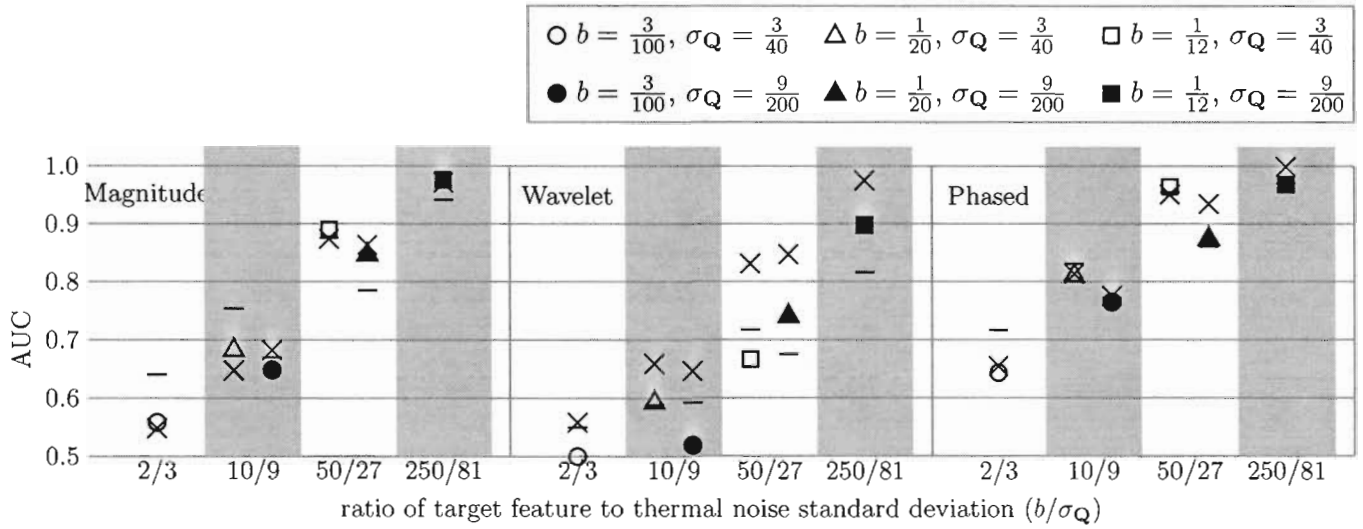

Figure 3.11: **Experiment Two**: Plot of the mean AUC of the the model and mean AUC of the 16 human observers in each of the 18 experimental conditions. The y-axis is the AUC score. Each experimental condition is represented by a symbol (one of three shapes, either filled or unfilled) that locates the mean human observer AUC. The symbol's shape represents a different configuration of target feature and noise power, as explained in the legend at the top-right of the chart. Either directly above or below the symbol will be one horizontal bar and one x-mark. The horizontal bar represents the mean NPWE result and the x-mark represents the CHO result for the experimental condition. The x-axis is divided into thirds, with each third containing results from one of the estimation techniques as labeled in the top-left corner of each third. Within each third, the x-axis is then further divided by the ratio of target feature to thermal noise standard deviation (complex feature-SNR). In some cases, two experimental conditions have the same PSNR (e.g,. $b = \frac{1}{12}, \sigma_Q = \frac{3}{40}$ and $b = \frac{1}{20}, \sigma_Q = \frac{9}{200}$) and so appear in the same band on the diagram.

discussing this, we will lay out some general ideas that may be helpful in choosing evaluation methods for MRI.

One of the principal goals of the experiments discussed above was to determine whether a best-case phase-corrected real reconstruction could produce noticeable differences between the images and, in particular, differences that improved feature detection. While we have not demonstrated statistically significant results — due to our small number of subjects relative to the number of conditions explored — our experiments do suggest that dim features in the dark parts of the image are more detectable in phase-corrected images than magnitude ones. Further, our experiments suggested that features in bright parts of the image are not more detectable with phase-corrected real reconstruction.

As we noted above, these results match predictions made based on the bias and MSE of the two estimators — phase-corrected real reconstructions are lower bias and have lower MSE than magnitude images in the dark parts of the image, but are essentially the same in bright parts. However, in the part of our experiment comparing wavelet reconstructions our results contradicted predictions based on 2-norm error measures. While these results may lead one to doubt the value of pixel 2-norm in gauging MRI estimator quality, if we return to the discussion in section 3.2 we can distinguish why pixel MSE predicts some of our results and not others.

We previously noted that RMSD and other 2-norm error measures are unlikely to be useful when the estimator uses regularization across pixels. This is because many transforms that we roughly defined as 'perceptually tricky' (blurring, shifts, etc.) can be produced by these estimators, and these changes are poorly captured by the 2-norm. The wavelet estimator that we studied is an estimator of this type – it uses neighbourhoods of each pixel's measurements to estimate the true pixel value – and so we should not expect 2-norm error measures to predict human feature detection performance with this estimator.

In contrast, the magnitude estimator is an independent estimator, and so 2-norm error measures are more reasonable in evaluating changes in quality between images of this type. In fact, the historical use of 2-norm error measures to evaluate the improvement in magnitude images produced by changes in hardware has likely influenced the widespread use of these measures in other aspects of the MRI community. Much of the early work on quantifying the quality of MRI measurements focused on comparisons between hardware, principally between different magnets and coil setups. Referring to the PDF in equation

(2.3) we can describe the effects that changes to these components should produce as a constant scaling of s at every pixel and change in the value of σ [35, 37, 27, 50, 48]. If these are the changes we expect to see between the images we are comparing, 2-norm error measures are indeed reasonable. Imaging systems with equal ratios of signal to noise variance are effectively equivalent, and this ratio will be captured in 2-norm error measures. These error measures, however, will not effectively quantify the difference between the variety of images that can be produced by regularized estimators. Thus, using these values to evaluate the quality of wavelet-thresholded MR images is probably ill-advised.

In chapter 4 we will consider independent estimators. Relying on the discussion and results of this chapter, we will use bias and MSE as the metrics for our preliminary comparison of these estimators. In particular, we will focus on the idea that, given two estimators, if one has lower MSE at all possible values of the true parameters, this estimator is likely to be preferable to the high-MSE estimator for feature detection.

However, in chapter 4 we are also interested in determining whether any of the independent estimators produce results similar to the ideal phase-corrected real reconstruction. Since we have previously stated that phase-corrected estimators are regularized estimators, this may seem to contradict the conclusion we just made about the use of 2-norm error measures and regularized estimators. We make this exception because the *best-case* phase-corrected real estimator is independent. As we stated in section 3.5.1, if the true phase ϕ is known, as we assume in the best-case estimator, then each pixel can be phase-corrected independently and there is no correlation introduced between neighbouring pixels. The use of regularization occurs in practical implementations of phase correction since ϕ is not known and so is estimated using a regularized estimator. However, since chapter 4 is only concerned with comparing practical, independent estimators with the idealized, independent, best-case phase-corrected estimator, our use of 2-norm error measures is reasonable.

3.7 Summary

In this chapter we have surveyed a variety of approaches for comparing medical image estimators. We have presented human experiments that we developed to evaluate MR image estimators and have shown that our results suggest the best-case phase-corrected real estimators do indeed provide improved detection compared to magnitude images. This leads to our work in chapters 4 and 5 where we will attempt to derive practical estimators

that implement these results. The human results were also used to provide the first results demonstrating the applicability of CHO and NPWE model observers to MRI data.

We have also discussed the applicability of different error measures to the comparison of different types of estimators. 2-norm error measures can be justified when applied to independent estimators, and their historical use in evaluating MRI hardware is quite reasonable. Although these error measures should not be used generally in the comparison of regularized estimators, we have justified our use of bias and MSE in the remainder of this thesis by noting that the phase-corrected real estimator is unusual in that its best-case formulation, which is the target of our comparisons, is an independent estimator while all practical implementations of phase correction are regularized estimators.

Chapter 4

Independent Estimation of Signal Magnitude

The experimental results shown in chapter 3 suggest that an estimator approximating the best-case phase-corrected real estimator can improve detection of some features. Based on these results, we were interested in determining whether an independent two-parameter estimator could produce a result approximating the desired performance. If such an independent estimator could be constructed, it would avoid the difficulties that previous attempts to regularize the phase parameter had encountered and provide the benefits of phase-corrected real reconstruction without the attendant drawbacks.

We begin this chapter with a brief survey of parametric point estimation theory. Following this survey, we introduce the previous work on independent estimators for MRI, most of which has focused on the one-parameter model. We then apply the statistical estimation theory we have surveyed to the two-parameter model of the MRI signal and relate the resulting estimators to those previously published. Finally, we compare our newly derived estimators with those previously published and with the best-case phase-corrected real estimator and draw some conclusions from the results.

4.1 Survey of Parametric Point Estimation Theory

Parametric estimation problems ask us to summarize our knowledge of some parameters given some measurements. Normally these parameters are thought of as the parameters of

some distribution from which the measurements are drawn. For example, given a set of measurements from $\mathcal{N}(\mu, \sigma)$ we might be asked to summarize our knowledge of μ and σ. There are, very generally, two major schools of thought in statistics about how this problem should be approached, which are dubbed *Frequentist* and *Bayesian*. Without going into details, in this work we will restrict ourself to what are normally considered frequentist methods of summarizing knowledge, although it is likely that identical results could be produced via Bayesian approaches with the assignment of different definitions and assumptions. The choice of frequentist methodology is not due to any preference for the philosophy, so much as the fact that it has been previously employed in MRI and so the relationship between old estimators and new estimators generated in the frequentist framework is likely easier to understand.

The parametric estimation problem we are most concerned with is summarizing our knowledge of the true signal magnitude, given measurements drawn from the distribution shown in equation (2.3). In the frequentist methodology, there are many ways that we can express our knowledge of the parameters given the measurements. One approach is to assign some range of values to the parameters along with a definition of what the range means (e.g., confidence intervals). However, if we assign a range of values to the true value of the signal magnitude at each voxel, it is unclear how we should represent this as an image. Instead of a single greyscale value, we would then have a range of possible values at each pixel. In order to produce a greyscale image we must assign a single value as our summary of knowledge about the true signal magnitude at each point. Such a summary is called a *point estimate*, and so we will now provide a brief survey of the theory of parametric point estimation.

4.1.1 Likelihood and maximum likelihood estimators

The notion of a *likelihood* is central to frequentist methodology for performing inference. A likelihood is a PDF viewed as a function of its parameters with its measurements fixed. It is important to note that a likelihood does not have many properties common to PDFs. For example, integrating a likelihood function over its parameters will not usually produce "1" as the answer. Similarly, a likelihood should not be confused with a probability of a parameter taking a certain value given the measurements. In frequentist statistics, one cannot assign a probability to parameters given measurements, and so it is critical to distinguish likelihoods (variable parameters, fixed measurements) from probabilities (variable measurements, fixed

parameters). We write the likelihood of the model with parameters α and β given the fixed vector of measurements $\boldsymbol{x}$ as $L(\alpha, \beta; \boldsymbol{x})$.

The likelihood allows us to assign a numerical value to any choice of point estimate for the parameters. The most common approach is to then pick the point estimate which maximizes the likelihood. This is called the maximum likelihood estimator (MLE). If we think of the PDF as describing error due to 'noise', as is the case in our model of the MRI signal, then maximizing the likelihood is equivalent to picking the point estimate which assumes the smallest amount of noise in the measurement. Noting that taking the logarithm of a function does not change its extrema, it is often more convenient to maximize the log of the likelihood function, which we will denote $\ell(\alpha, \beta; \boldsymbol{x})$. This is so common, in fact, that there is a name, *score*, for the gradient of the log-likelihood, which we will denote $U(\alpha, \beta; \boldsymbol{x})$. We will use U_α to denote the score in the direction of parameter α, and U_β to denote the score in the direction of β.

Setting $U(\alpha, \beta; \boldsymbol{x}) = 0$ and solving for the parameters gives the extreme points of the likelihood, and second or higher derivatives of $\ell(\alpha, \beta; \boldsymbol{x})$ can be used to select point estimates that are maximal. It is generally the case that most 'well-behaved' likelihoods will have only one maximum, and we will find that this generally holds for our problem.

4.1.2 The Fisher information matrix and the Cramér-Rao bound

As we noted above, the second derivative of the log-likelihood is useful for helping determine maximum likelihood points. However, they also are part of the definition of the Fisher information matrix $\mathbf{I}$, which is defined as the negative expected value of the matrix of all possible second derivatives of ℓ with respect to the parameters. This matrix can be thought of as the covariance matrix of the individual components of the score function.

Equally important to our work is the Cramér-Rao bound derived from the information matrix. The basic definition of the bound is that the inverse of the Fisher information matrix provides a lower bound on the variance of any unbiased estimator of the parameters. Thus, if we desire an unbiased estimator of our MRI signal parameters, the Cramér-Rao bound determines what amount of variance we must be prepared to accept in our result. It is important to note that it does not guarantee that there is an unbiased estimator that achieves such variance. However, any unbiased estimator that does achieve the bounding variance is said to be *efficient*.

The Cramér-Rao bound is used in many problems as part of the motivation for choosing

the maximum likelihood estimator over any other possible choice. This is because it can be shown that when the Fisher information matrix is non-zero and well-defined, the MLE is asymptotically unbiased and efficient as the number of measurements increases to infinity [8, 41]. Of course, in problems such as ours where the number of measurements is small — in many cases there is only one measurement per location — these optimality criteria for the MLE are not sufficient.

4.1.3 Reduction by Sufficiency

It is common to have a PDF specified that describes one measurement from a system, and then take multiple independent measurements as part of an experiment and try to make a single estimate of the parameters using all the measurements. One way of doing this is to produce a joint PDF for all the measurements, and then treat this as a joint likelihood from which the parameters can be inferred. Since the measurements are independent, the joint PDF is the product of the individual PDFs. However, this can quickly become cumbersome particularly if one is trying to derive estimators for the general case of n measurements where n might be different for any individual experiment.

An alternative approach to use a single quantity, computed from all the measurements, as a summary of all the measurements. When a reduction like this is possible, the reduced quantity is called a *sufficient statistic*. The most common way of proving that a given function of the data $q(\boldsymbol{x})$, is a sufficient statistic for n measurements $\boldsymbol{x} = \{x_1, \ldots x_n\}$ is to show that the joint PDF $p(\boldsymbol{x}; \alpha)$ can be factored as

$$p(\boldsymbol{x}; \alpha) = f(\boldsymbol{x}) g(q(\boldsymbol{x}), \alpha)\,, \tag{4.1}$$

where f and g are arbitrary functions. In this case we can use the PDF $p(q(\boldsymbol{x}); \alpha)$ for estimation of α without loss of information [41].

4.1.4 Nuisance parameters and likelihood-like functions

It is often the case that, although a likelihood is specified with multiple parameters, only one of the parameters is actually of interest for the problem at hand. In our problem, we can see that the parameter s is the parameter of interest, while ϕ and σ are simply nuisance parameters. We are indifferent to the point estimate of ϕ and σ because only the point estimate of s affects the image we will generate. The problem of estimating only one

parameter from a model with many parameters is a significant problem in statistics and we will now survey some of the suggested approaches that we will use in later sections. Since none of the models to which we will apply these approaches contain more than one parameter of interest and one nuisance parameter, we will restrict our survey to models of this type, even though the theory presented generally extends to any number of parameters of interest and nuisance parameters.

Profile likelihood

Perhaps the most common intuition about how to solve this problem is to simply take the maximum likelihood estimate for all the parameters and ignore those that are not of interest. More formally, let us define a likelihood-like function called the *profile likelihood*. Given a likelihood $L(\alpha, \beta; \boldsymbol{x})$ where α is a parameter of interest, β is a vector of nuisance parameters, and $\boldsymbol{x}$ a vector of measurements, we define the profile likelihood

$$L_{\mathrm{P}}(\alpha) = \max_{\beta} \ell(\alpha, \beta; \boldsymbol{x}) \,, \tag{4.2}$$

where $\max_{\beta} f(\beta)$ is the maximum value of f achieved over the range of valid values for β. Taking the logarithms of both sides gives us the definition of the profile log-likelihood ℓ_{P} and taking the derivative with respect to α gives the profile score. While $\ell_{\mathrm{P}}(\alpha; \boldsymbol{x})$ is generally not the same as $\ell(\alpha, \beta; \boldsymbol{x})$, it is true that maximizing $\ell_{\mathrm{P}}(\alpha; \boldsymbol{x})$ produces the same estimate of α as maximizing the full likelihood [8].

Maximum marginal likelihood estimate

Given a PDF of the form $p(\boldsymbol{x}, \boldsymbol{y}; \alpha, \beta)$, with measurements divided into two vectors $\boldsymbol{x}, \boldsymbol{y}$, parameter of interest α, and nuisance parameter β, a marginal likelihood for α can be constructed whenever marginalizing out the measurements $\boldsymbol{y}$ will produce

$$p(\boldsymbol{x}; \alpha) = \int p(\boldsymbol{x}, \boldsymbol{y}; \alpha, \beta) \, \mathrm{d}\boldsymbol{y} \,. \tag{4.3}$$

In other words, marginalization can be applied when integrating over the values of $\boldsymbol{y}$ simultaneously removes the dependence on the nuisance parameter [36, 11]. This is sometimes referred to as $\boldsymbol{x}$ being α-oriented [11]. By the *general sufficiency principle* this reduction is only information-preserving when $\boldsymbol{x}$ is α-oriented and sufficient for α when β is held fixed [11].

Maximum Bartlett-corrected profile likelihood

A variety of modifications to the profile likelihood have been suggested to improve its performance. We will first consider Bartlett's modification to the profile score [10, 42]. Bartlett uses the first-order model of the full likelihood's score. In this approximation, the derivative with respect to each parameter is assumed to be normally distributed with zero mean and covariance given by the Fisher information matrix. If we denote the Fisher information matrix as $\mathbf{I}$ and split it (and its inverse) into quadrants addressed as (α, α), $(\alpha, \boldsymbol{\beta})$, or $(\boldsymbol{\beta}, \boldsymbol{\beta})$ (only three are needed since $\mathbf{I}$ is symmetric), we can write the approximation to the score function as

$$U(\alpha, \beta; \boldsymbol{x}) \simeq U_\alpha(\alpha, \beta; \boldsymbol{x}) - cU_\beta(\alpha, \beta; \boldsymbol{x}) \tag{4.4}$$

where $c = \mathbf{I}_{\alpha,\beta}\mathbf{I}^{-1}_{\beta,\beta}$. We note that if $\mathbf{I}_{\alpha,\beta} = 0$, the second term of this model must be zero. When $\mathbf{I}_{\alpha,\beta} = 0$ we say that the nuisance parameters are orthogonal to the parameter of interest. Fortunately, because α is a scalar, there is always a transformation of the parameters that makes the nuisance parameters orthogonal to the parameter of interest [8]. Thus, if the parameters are not orthogonal in the original specification, we can transform β to a new nuisance parameter ζ that is orthogonal to α. In this way, we can write our model for the score function using only one term. We also note that this transformation does not affect $\ell_{\mathrm{P}}(\alpha; \boldsymbol{x})$ [8]. Thus, a correction determined with either parameterization can be applied to $\ell_{\mathrm{P}}(\alpha; \boldsymbol{x})$.

If we perform a change of parameters such that the nuisance parameters are orthogonal to the parameter of interest, we can write Bartlett's proposed correction as

$$\Delta = -\frac{1}{2}\mathbf{I}'^{-1}_{\zeta,\zeta} E\left(\frac{\partial^3 \ell'(\alpha, \zeta)}{\partial\alpha\,\partial\zeta\,\partial\zeta}\right) \tag{4.5}$$

where $\mathbf{I}'$ is the information matrix for the transformed parameters and $\ell'(\alpha, \zeta; \boldsymbol{x})$ is the transformed log likelihood. This correction is then subtracted from the profile score $\frac{\mathrm{d}\ell_{\mathrm{P}}(\alpha;\boldsymbol{x})}{\mathrm{d}\alpha}$ to produce the Bartlett-corrected profile score. This new function can be treated as a score and thus set equal to zero and solved for the maximum Bartlett-corrected profile likelihood estimator.

Maximum stably adjusted profile likelihood estimate

The stably adjusted profile likelihood approximates the likelihood that would result if nuisance parameters could be removed in an information-preserving factorization. It does this

by calculating a weight $M(\alpha)$ such that $M(\alpha)L_{\mathrm{P}}(\alpha;\boldsymbol{x})$ is the approximation and can be maximized to produce an estimate, just like a standard likelihood. If α is the parameter of interest, and β is a nuisance parameter, we begin as in the Bartlett-corrected case by determining a change of parameters such that the nuisance parameter ζ is orthogonal to α. We will similarly use primes (e.g., $\mathbf{I}'$) to represent reparameterized functions of the model. The stably adjusted profile likelihood is given by the weighting [8, 7]

$$M(\alpha) = \sqrt{\frac{\mathbf{I}'_{\zeta,\zeta}}{J}}\Bigg|_{\zeta=\zeta_\alpha} \exp(g(\alpha))\,, \tag{4.6}$$

where $J = -\frac{\partial^2}{\partial\zeta^2}\ell'(\alpha,\zeta;\boldsymbol{x})$, ζ_α is the solution of $U_\zeta(\alpha,\zeta;\boldsymbol{x}) = 0$ in terms of α, and

$$g(\alpha) = \int_{\hat{\alpha}_{\mathrm{ML}}}^{\alpha} \left\{ \mathbf{I}'^{-1}_{\zeta,\zeta} \left[\frac{1}{2} E\left(\left(\frac{\partial \ell'(\alpha,\zeta;\boldsymbol{x})}{\partial\zeta} \right)^2 \frac{\partial \ell'(\alpha,\zeta;\boldsymbol{x})}{\partial\alpha} \right) + E\left(\frac{\partial^2 \ell'(\alpha,\zeta;\boldsymbol{x})}{\partial\zeta^2} \frac{\partial \ell'(\alpha,\zeta;\boldsymbol{x})}{\partial\alpha} \right) \right] \right\} \Bigg|_{\zeta=\zeta_\alpha,\,\alpha=z} \mathrm{d}z \tag{4.7}$$

4.2 Formulation of the Signal Magnitude Estimation Problem

In sections 2.1 and 2.2 we provided the PDFs for the full complex measurement — equation (2.3) — and the magnitude of this complex measurement — equation (2.4). In this section we will now use these equations to formalize the two major formulations of the signal magnitude estimation problem. We begin with what we call the two-parameter model and then show how it can be used to derive the one-parameter model.

However, we first need to clarify exactly what measurements we are performing estimation from and how they are obtained. For our purposes, we will assume that the complex measurement at each voxel has been repeated n times, where n is the same for all voxels. This sort of repeated measurement is usually achieved by repetition of the pulse sequence (the value of n is often called *number of excitations* or NEX on MRI scanners). Each of these measurements is independently drawn from the same distribution. Additionally, the samples are normally repeated quickly enough that registration does not become an issue. In the next two subsections, we will discuss different distributions that can be used to model these measurements and how they define the point estimation problem of turning the n measurements at each point into a single estimate of s.

The other significant feature of our measurements is that σ is fixed at a common value for all voxels, while the other parameters are free to assume different values at every voxel. The knowledge that σ is fixed is particularly useful when combined with the fact that almost all MR images include regions of air. Since we know that there is no signal emitted in voxels containing air, we can use these voxels to estimate the only remaining parameter of the PDF: σ. Then, since σ is the same at all voxels, we can extrapolate this estimate to all the other voxels and treat σ as known. If $n > 1$, we can also use the difference between all the measurements at each voxel to estimate σ even more robustly. Whichever method we choose, it is usual to assume that σ is known [34, 50, 55, 32]. The one significant exception to this, that we know of, is the independent estimator derived by Koay *et al.*, that assumes σ can vary between voxels and attempts to solve for this parameter as well [38]. For our purposes, a model where σ is common to all voxels is sufficient, and so we will proceed with this assumption.

4.2.1 Two-parameter model

The PDF shown in equation (2.3) has three parameters: s, ϕ, and σ. However, as we mentioned above, we can treat σ as known. Thus at every voxel, we have a model with two free parameters: s is our parameter of interest and ϕ is a nuisance parameter. We will write this reduced model as $p(r, \theta; s, \phi)$ and define it as as the PDF given in equation (2.3) with σ treated as a constant.

In order to apply any of the point estimation techniques presented above, we need a single PDF that can be turned into a likelihood. However, we have n measurements, and so $p(r, \theta; s, \phi)$ cannot be directly applied. As suggested in section 4.1.3, we will apply reduction by sufficiency to our measurements. Noting that our two-parameter model is a binormal distribution with independent measurements in the real and imaginary channels, the optimal sufficient statistic is the complex mean of the complex measurements [41]. The PDF of the mean measurement is exactly the same as $p(r, \theta; s, \phi, \sigma)$, although the standard deviation is now σ/n. Of course, since we are estimating σ from the data, we can perform this estimation using the complex mean data, and then the factor of $1/n$ will be taken into account automatically. Thus, if we immediately take the complex average of all the measurements at each voxel, we can then proceed to perform estimation as if we had only taken one measurement.

Having reduced the data to a single complex measurement at each voxel, and explained

a method for removing the parameter σ from the model, the PDF of this model is

$$p(r, \theta; s, \phi) = \frac{r}{2\pi\sigma^2} \exp\left(-\frac{s^2 + r^2 - 2sr\cos(\theta - \phi)}{2\sigma^2}\right), \tag{4.8}$$

which is essentially the same as equation (2.3) except now we let (r, θ) be the polar coordinates of the average complex measurement and σ is treated as a constant.

From this PDF we can now derive the Fisher information matrix for this problem, which we will denote $\mathbf{I}_2$ to make clear it describes the two-parameter model. Ordering the parameters s, ϕ, the matrix is then

$$\mathbf{I}_2 = \begin{bmatrix} \frac{1}{\sigma^2} & 0 \\ 0 & \frac{s^2}{\sigma^2} \end{bmatrix}. \tag{4.9}$$

The most important choice remaining in constructing estimators from the two-parameter model is how to deal with the existence of the nuisance parameter. Although some work has considered this question [71], to the best of our knowledge there has been no previous systematic application of the statistical theory to this problem. We have published the first results in this area [82], and this chapter extends these published results still further. We will consider a variety of derivations in section 4.4.

4.2.2 One-parameter model

The PDF of just r with θ discarded was previously given in equation (2.4), and this forms the basis of what we call the one-parameter model. The choice of this model as a topic of previous research seems to be based on the understanding that normally θ is discarded by MRI scanner software and only r is reported. Thus, an estimator for the two-parameter model that required measurements of θ would not be useful without being integrated into the scanner. The implied connection in past work is that, having discarded the measurements of θ, θ should also be marginalized out of the model that is used for inference. The fact that this reduction removes the dependence on the parameter ϕ seems to be treated as simply a fortuitous by-product and not the goal of the marginalization. We will return to these ideas in section 4.6.

When using the one parameter model, it is still possible to estimate σ from the data, and thus we write the PDF used for estimation as $p(r; s)$, treating σ as known. As in the two-parameter case, we have n measurements and would like to perform a reduction by sufficiency. One common assumption is that the complex measurements were averaged by

the scanner before the phase data was discarded. Thus, even with n measurements, r is taken to be the single real value representing the magnitude of the complex average. This is what is commonly called the *magnitude image*.

We note that it has been suggested anecdotally to the author that some MRI systems define the magnitude image as the mean of the magnitude of the n complex measurements. However, the average magnitude is not a sufficient statistic for the n magnitudes and doing such a reduction would discard data that could be used for inference. Regardless, since we lack a clear description of how any given scanner constructs its magnitude image, in this work we proceed with the definition that magnitude images are produced by averaging the complex measurements and then taking the magnitude of the average. Using this assumption the PDF for the one-parameter model at each voxel is then

$$\begin{aligned} p(r;s) &= \int_{-\pi}^{\pi} p(r,\theta;s,\phi)\,\mathrm{d}\theta \\ &= \frac{r}{\sigma^2}\exp\left(-\frac{r^2+s^2}{2\sigma^2}\right)\mathrm{I}_0\left(\frac{sr}{\sigma^2}\right) \end{aligned} \tag{4.10}$$

where r is the magnitude of the average complex measurement and σ is treated as a constant.

Unlike the two-parameter case, there is no nuisance parameter to be dealt with, so the inference process is clearly defined in frequentist statistical theory. Additionally, the Fisher information matrix for this model has only one entry, which we will denote $\mathbf{I}_1$ for the sake of consistency with our previous notation. Following our assumptions about how data are reduced, we have [71]

$$\mathbf{I}_1 = \frac{1}{\sigma^2}\left[E\left(\frac{r^2}{\sigma^2}\left[\frac{\mathrm{I}_1\left(\frac{sr}{\sigma^2}\right)}{\mathrm{I}_0\left(\frac{sr}{\sigma^2}\right)}\right]^2\right) - \frac{s^2}{\sigma^2}\right]. \tag{4.11}$$

In the next section, we will look at some estimators that have been developed by other authors using the one-parameter model.

4.3 Previously Published Independent Estimators for Signal Magnitude

Currently, most clinical diagnoses are performed using magnitude images. In these images, the estimated true signal magnitude s is given by

$$\widehat{s}_{\mathrm{Mag}} = r\,. \tag{4.12}$$

As we stated before, this estimator is known to have a Rician distribution [32, 48] with bias and mean squared error (MSE)

$$E(\widehat{s}_{\text{Mag}} - s) = \sigma\sqrt{\frac{\pi}{2}}{}_1\text{F}_1\left(-\frac{1}{2}; 1; -\frac{s^2}{2\sigma^2}\right) - s \tag{4.13}$$

$$E\left((\widehat{s}_{\text{Mag}} - s)^2\right) = 2(s^2 + \sigma^2) - s\sigma\sqrt{2\pi}{}_1\text{F}_1\left(-\frac{1}{2}; 1; -\frac{s^2}{2\sigma^2}\right), \tag{4.14}$$

where ${}_1\text{F}_1$ is the confluent hypergeometric function.

A variety of other estimators have been proposed for the independent estimation of the true signal magnitude from the one-parameter model. The great majority of these [32, 34, 50, 55, 38] have been based on the observation that

$$E[r^2] = s^2 + 2\sigma^2\,. \tag{4.15}$$

This result is a consequence of r^2/σ^2 having a non-central χ^2 distribution with two degrees of freedom. McGibney and Smith [50] and Miller and Jospeh [55] independently proposed the same estimator, which will call the MM estimator for brevity. This estimator was created by replacing $E[r^2]$ with r^2 in this equation, and then solving for s, giving

$$\hat{s}_{\text{MM}} = \sqrt{\widehat{s}^2_{\text{Mag}} - 2\sigma^2}\,. \tag{4.16}$$

Put another way, they find an unbiased estimate of s^2 and then take the square root of this quantity as their estimate of s. Note that taking the square root means that $\hat{s}_{\text{MM}}$ is not an unbiased estimate of s, even though $\hat{s}^2_{\text{MM}}$ is an unbiased estimate of s^2.

One complaint that has been leveled against this estimator is that it can produce imaginary-valued estimates for the real-valued s. It has been suggested that this means $\hat{s}_{\text{MM}}$ "cannot be a valid estimator" when s/σ is small [72]. We have a different interpretation of this that leads to a particular solution to the imaginary estimates. $\widehat{s}^2_{\text{Mag}} - 2\sigma^2$ is an unbiased estimator of s^2 and so, when $s \simeq 0$, the unbiased estimator must be able to take negative values. However, since we know that s^2 must be positive, there is another estimator that can be derived from $\widehat{s}^2_{\text{Mag}} - 2\sigma^2$ that is guaranteed to have equal or smaller error: $\max(0, \widehat{s}^2_{\text{Mag}} - 2\sigma^2)$. Taking the square root then gives us

$$\sqrt{\max(0, \widehat{s}^2_{\text{Mag}} - 2\sigma^2)} = \text{Re}(\sqrt{\widehat{s}^2_{\text{Mag}} - 2\sigma^2})\,. \tag{4.17}$$

Thus, by simply taking the real part of $\hat{s}_{\text{MM}}$ (which is likely what MacGibney and Smith, and Miller and Joseph intended, although neither paper is explicit in this [50, 55]) we are basing our result on an improved estimator of s^2 and producing a valid estimator.

The estimator proposed by Gudbjartsson and Patz, which we will call the GP estimator, is quite similar to this interpretation of $\hat{s}_{\mathrm{MM}}$. Starting with just the Rician-distributed magnitude measurements, they propose to make the resulting estimator's distribution closer to Gaussian by using [32]

$$\hat{s}_{\mathrm{GP}} = \sqrt{\left|\hat{s}^2_{\mathrm{Mag}} - \sigma^2\right|}\,. \tag{4.18}$$

Note the differences between this and equation (4.16) are the introduction of the absolute value inside the square root and a change in the coefficient of the σ term.

An alternative to these approaches is to begin with the one-parameter model and perform maximum likelihood estimation. An analysis of this idea has previously been published in the context of astronomy by Simmons and Stewart [73, 75] and then independently reanalyzed by Sijbers *et al.* more recently in the context of MRI [72, 71]. The work by Sijbers *et al.* provides a more thorough description of how they arrive at their results and presents a clear algorithm for performing maximum likelihood estimation on the one-parameter model. The basis of this model begins by setting the one-parameter model's score function equal to zero

$$U(s;r) = \frac{1}{\sigma^2}\left(r\frac{\mathrm{I}_1\left(\frac{sr}{\sigma^2}\right)}{\mathrm{I}_0\left(\frac{sr}{\sigma^2}\right)} - s\right) = 0\,. \tag{4.19}$$

The solutions of this equation were then studied using catastrophe theory [72]. The basic result of this analysis is that the one-parameter maximum likelihood estimate of s is 0 when $r^2 \leq 2n\sigma^2$. Otherwise, there is one positive maximum which can be found numerically. We implemented the numerical root finding using Mathematica's implementation of Brent's method [90]. For brevity, we will call this the Sijbers estimator.

4.4 Our Derivations of Independent Two-Parameter Estimators for MRI

In this section we will apply the statistical theory we surveyed in section 4.1 to the two parameter model to derive a variety of estimators for the MRI signal magnitude. We will find that some of the estimators derived using these methods are the same as previous estimators. When this overlap occurs, we will highlight how the previously published estimator relates to the statistical theory.

4.4.1 Maximum likelihood estimate

Let $(\hat{s}_{\mathrm{ML}}, \hat{\phi}_{\mathrm{ML}})$ be the MLE of (s, ϕ) computed by solving

$$U(s, \phi; r, \theta) = 0\,, \tag{4.20}$$

which produces the two equations

$$-\frac{s - r\cos(\theta - \phi)}{\sigma^2} = 0 \tag{4.21}$$

$$\frac{rs\sin(\theta - \phi)}{\sigma^2} = 0\,. \tag{4.22}$$

The solutions that maximize the likelihood are

$$\hat{s}_{\mathrm{ML}} = r\cos(\theta - \hat{\phi}) \tag{4.23}$$

$$\hat{\phi}_{\mathrm{ML}} = \theta + n\pi\,, n \in \mathbb{Z}\,. \tag{4.24}$$

The choice of n even or odd selects the sign of $\hat{s}_{\mathrm{ML}}$. However, since we know that $s \geq 0$, it is reasonable to select the positive version of $\hat{s}_{\mathrm{ML}}$ as the maximum likelihood estimate. Noting that $\hat{s}_{\mathrm{ML}}$ is just the magnitude image, we know that this estimator has a Rician distribution with bias and MSE as given in equations (4.13) and (4.14).

Although we know that the maximum profile likelihood will give the same answer, the profile likelihood function will be useful later so we will derive it here. We do this by noting that $\phi = \theta$ is the solution to $U_\phi(s, \phi; r, \theta) = 0$ and substituting this into the likelihood to produce

$$L_{\mathrm{P}}(s; r, \theta) = \frac{r}{2\pi\sigma^2}\exp\left(-\frac{(r-s)^2}{2\sigma^2}\right)\,. \tag{4.25}$$

Although this result is somewhat trivial mathematically, it is important because previous authors have referred to the one-parameter MLE as "the MLE" of the MRI signal. In reality, the magnitude image is the MLE of the two-parameter MRI model. In the next subsection we will determine the relationship of the one-parameter MLE to the two-parameter model.

4.4.2 Maximum marginal likelihood estimate

In the two-parameter model we have proposed, marginalizing out θ gives the one-parameter model

$$p(r; s) = \frac{r}{\sigma^2}\exp\left(-\frac{s^2 + r^2}{2\sigma^2}\right)\mathrm{I}_0\left(\frac{sr}{\sigma^2}\right)\,. \tag{4.26}$$

As we noted above, maximizing the one-parameter model's likelihood has previously been presented by Sijbers *et al.* [72, 71]. However, our interest here is understanding whether this reduction meets the criteria for being information-preserving that we laid out in section 4.1.4.

Since this marginalization removes the dependence on ϕ, we can see that r is s-oriented [11]. The remaining problem in determining whether the marginalization is information preserving is to test if r is sufficient for s when ϕ is held constant in the two-parameter model. This is equivalent to finding a factorization of [41]

$$p(r, \theta; s, \phi) = f(r, \theta, \phi)g(r, s, \phi)\,. \tag{4.27}$$

However, consulting equation (4.8), we can see that such a factorization does not exist due to the existence of the factor $\exp\left(\frac{sr\cos(\theta-\phi)}{2\sigma^2}\right)$ in $p(r, \theta; s, \phi)$. Thus, we should not expect this marginalization to be information preserving and it would not be recommended as an inference approach under the general sufficiency principle [11].

Regardless of the information-preserving properties of the reduction, we will consider this marginalization because it describes an alternate methodology that can be used to produce the Sijbers estimator [72, 70, 71]. Interestingly, while our understanding is that the one-parameter model was used by Sijbers *et al.* based on its avoidance of θ measurements, we can equally interpret it, as we do here, in terms of parameter reduction. However, regardless of how the marginalization is justified, the loss of information suggested by the general sufficiency principle is problematic, and we will address this further in section 4.6.

4.4.3 Maximum Bartlett-corrected profile likelihood estimate

In section 4.1.4 we showed how to calculate Bartlett's correction for orthogonal nuisance parameters, and also how to convert any model to one with orthogonal nuisance parameters. However, looking at the information matrix for the two-parameter model in equation (4.9) we can see that our nuisance parameter is already orthogonal. Thus, we can immediately apply the simplified formula for the correction. Using the information matrix we compute the Bartlett correction for our problem as

$$\begin{aligned} \Delta &= -\frac{\sigma^2}{2s^2} E\left(\frac{\partial}{\partial s}\frac{\partial^2}{\partial \phi^2}\ell(s, \phi; r, \theta)\right) \\ &= \frac{1}{2s}\,. \end{aligned} \tag{4.28}$$

We can then use this to produce the corrected profile score function using the profile score function (the derivative of the log of $L_{\mathrm{P}}(s;r,\theta)$ as defined in equation (4.25))

$$U_{\mathrm{P}}(s;r,\theta) - \Delta = \frac{r-s}{\sigma^2} - \frac{1}{2s}\,. \tag{4.29}$$

Setting the corrected profile score function to zero and solving for s gives the maximum corrected profile likelihood estimate

$$\hat{s}_{\mathrm{Corr}} = \frac{\hat{s}_{\mathrm{ML}} + \sqrt{\hat{s}_{\mathrm{ML}}^2 - 2\sigma^2}}{2}\,, \tag{4.30}$$

where $\hat{s}_{\mathrm{ML}}$ is the MLE as described in section 4.4.1.

Comparing this estimator to equation (4.16), we can see that our new estimator is the average of $\hat{s}_{\mathrm{ML}}$ and $\hat{s}_{\mathrm{MM}}$. This means that, like $\hat{s}_{\mathrm{MM}}$, the Bartlett-corrected estimator can produce complex-valued estimates of the real-valued s. We will resolve this problem the same way we did with $\hat{s}_{\mathrm{MM}}$, by taking the real part of the complex-valued estimated.

4.4.4 Maximum stably adjusted profile likelihood estimate

As we noted in deriving Bartlett's correction, our two-parameter model's nuisance parameter is already orthogonal so we can directly apply the equation for the adjustment factor to parameters s and ϕ. We compute

$$J = \frac{rs\cos(\theta-\phi)}{\sigma^2} \tag{4.31}$$

$$\phi_s = \theta\,, \tag{4.32}$$

and

$$E\left(\frac{\partial^2\ell(s,\phi;r,\theta)}{\partial\phi^2}\frac{\partial\ell(s,\phi;r,\theta)}{\partial s}\right) = -\frac{s}{\sigma^2} \tag{4.33}$$

$$E\left[\left(\frac{\partial\ell(s,\phi;r,\theta)}{\partial\phi}\right)^2\frac{\partial\ell(s,\phi;r,\theta)}{\partial s}\right] = 0\,. \tag{4.34}$$

Substituting these equations into equation (4.7), and simplifying we can write

$$g(s) = -\int_r^s \frac{1}{z}\,\mathrm{d}z\,, \tag{4.35}$$

which simplifies to

$$g(s) = -\log(s/r)\,. \tag{4.36}$$

Using this identity, we can write the weighting factor, derived from equation (4.6), for this model as

$$M(s) = \sqrt{\frac{s}{r}\frac{r}{s}} = \sqrt{\frac{r}{s}}. \tag{4.37}$$

Combining this with the profile likelihood function gives the stably adjusted profile likelihood

$$L_{\mathrm{A}}(s; r, \theta) = (r/s)\frac{r}{2\pi\sigma^2}\exp\left(-\frac{(r-s)^2}{2\sigma^2}\right). \tag{4.38}$$

Taking the log and then maximizing we find that our maximum adjusted likelihood estimate is

$$\hat{s}_{\mathrm{A}} = \frac{\hat{s}_{\mathrm{ML}} + \sqrt{\hat{s}_{\mathrm{ML}}^2 - 4\sigma^2}}{2}. \tag{4.39}$$

Thus, the difference between the Bartlett-corrected estimate and the stably adjusted one is the coefficient of σ in the second term. This estimator can also produce complex-valued estimates and, using the same reasoning as before, we will use only the real part as our estimate.

4.5 Comparison of New and Previously Published Estimators

Having presented four previous estimators (magnitude image/maximum likelihood, MM, GP, and Sijbers/maximum marginal likelihood) and derived two new new ones (maximum Bartlett-corrected profile likelihood and maximum stably adjusted profile likelihood), we would like to compare these estimators with each other and the best-case phase-corrected real estimator. As discussed in section 3.6, we will use bias and MSE to analyze the performance of these estimators, since all of them are independent. We have presented a similar analysis previously with early results of this work [82], but this section extends the results with new estimators derived using the more thorough analysis of statistical theory laid out earlier in this chapter.

Based on the design of the best-case phase-corrected real estimator — as implemented in equations (3.9) and (3.10) — we know that this estimator has distribution $\mathcal{N}(s, \sigma)$ and thus is unbiased and has MSE σ^2. Since this estimator is unbiased we can compare it to the Cramér-Rao bound of the two-parameter model. This comparison shows that the best-case phase-corrected real estimator is the efficient estimator for the two-parameter model.

However, it is important to note that this distribution does not actually describe the images that we tested in chapter 3, because negative estimates at each pixel were set to zero, truncating the range of possible estimates. Thus, the bias and MSE of the estimator that produced our images are given by

$$E(\hat{s}-s) = \frac{\sigma}{\sqrt{2\pi}}\exp\left(-\frac{s^2}{2\sigma^2}\right) - \frac{s}{2}\left(1-\mathrm{erf}\left(\frac{s}{\sigma\sqrt{2}}\right)\right) \quad (4.40)$$

$$E\left((\hat{s}-s)^2\right) = \sigma^2 + \frac{(s^2-\sigma^2)}{2}\left(1-\mathrm{erf}\left(\frac{s}{\sigma\sqrt{2}}\right)\right) - \frac{2\sigma}{\sqrt{2\pi}}\exp\left(-\frac{s^2}{2\sigma^2}\right) \quad (4.41)$$

where erf is the error function. Clearly this is not an efficient estimator since it is biased. However, this is the estimator that we tested in our perceptual experiments, and so these bias and MSE values will be the baseline for comparison with our results below.

To compute the bias and MSE of all the estimators, we have used closed forms where possible and otherwise have estimated these values using a Monte Carlo simulation. At each data point our Monte Carlo simulation used 50,000 samples drawn from the two-parameter model and then processed as per the estimators. We ran the experiment from $s = 0$ to $s = 4$ at intervals of 0.25. In all of our simulations we used $\sigma = 1$, since it is the ratio of s to σ that controls the performance of these estimators and so altering σ would be equivalent to scaling s.

While useful for quantifying estimator performance in some cases, bias and MSE measurements do not lend themselves to qualitative descriptions of the resulting images. Thus, before proceeding with the quantitative analysis of the estimators, we present an example data set in figure 4.1 that may allow a more qualitative comparison of the estimators' output. In particular, we have set $\sigma = 1$ and used targets whose signal amplitude matches the range of values considered in the bias and MSE analysis that makes up the rest of this section.

4.5.1 Bias Results and Discussion

The bias of the estimators is plotted in figure 4.2. Comparing the bias of individual estimators (black curves) to the bias of the base-case phase-corrected images (grey curves), we see that none of the estimators we have presented in this chapter have bias performance similar to the phase-corrected estimator's bias. From our definition in section 4.1.2 we know that an efficient estimator would have zero bias at all values of s. Although none of our estimators

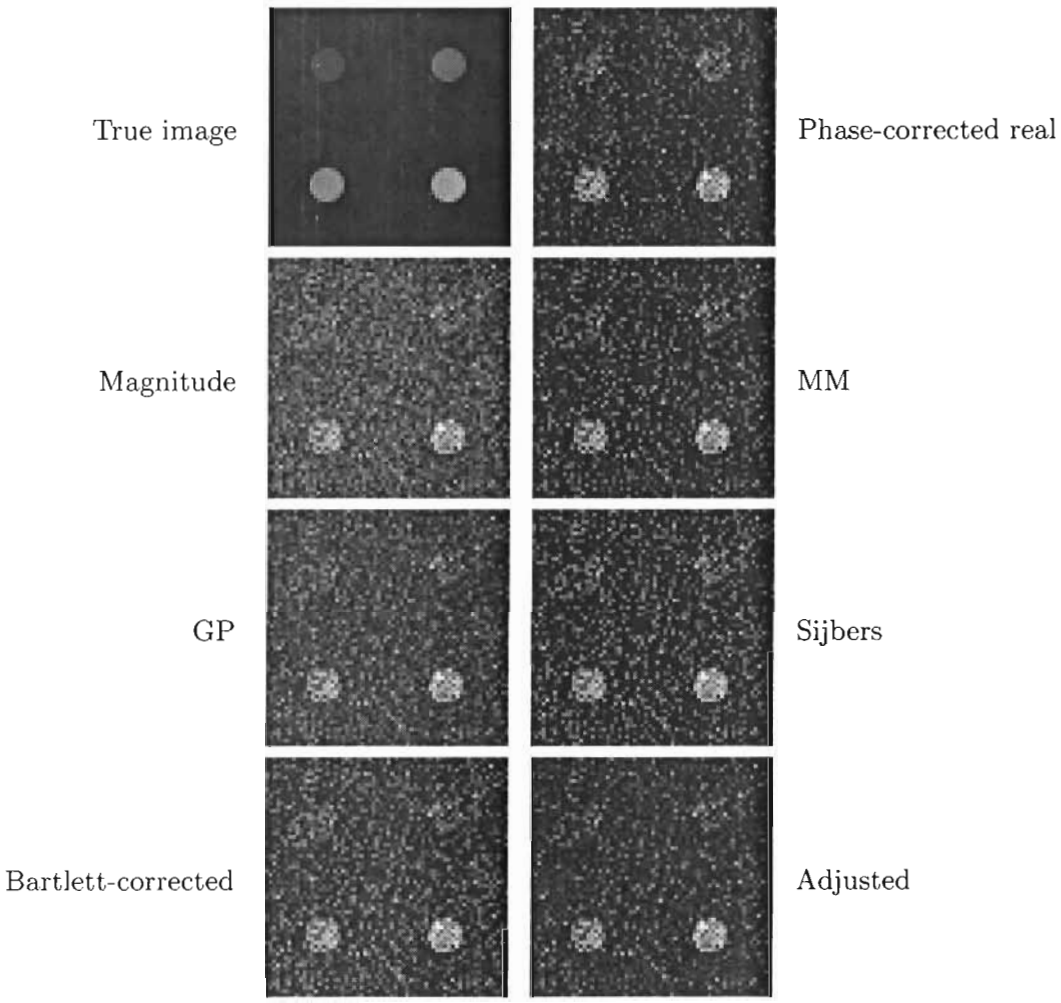

Figure 4.1: Example of independent estimator output. Complex zero-mean Gaussian noise with $\sigma = 1$ was added to the true image and the resulting magnitude data was used to produce estimated images. The four target circles have integer amplitudes from 1–4. The images are all scaled to the same units to enable correct comparison. The top row shows idealized reconstructions that are not possible with independent estimators. The remaining images show estimators discussed in this chapter.

match this performance, it is clear that the best-case phase-corrected images most closely approximate this behaviour.

Of course, the best-case phase-corrected image estimator is not a practical estimator since it requires prior knowledge of the true phase ϕ at every voxel. If we would like to avoid regularization completely, then we should ignore the best-case phase-corrected estimator and compare only the estimators that we have presented in this chapter. All the black curves in figure 4.2 represent estimators that can be practically implemented, and we can attempt to determine which is preferable among the practical independent estimators.

Comparing the black bias curves we can see that there are several trade-offs that warrant further evaluation. For example, the Sijbers/maximum marginal likelihood estimator (d in figure 4.2) has lower initial bias than most other estimators, but is slower to reach unbiasedness than either the maximum Bartlett-corrected profile likelihood or the GP estimators (e and c respectively in figure 4.2). Similarly, comparing these last two estimators, we see that the maximum Bartlett-corrected profile likelihood estimator has lower bias everywhere, but the GP estimator becomes unbiased slightly sooner. Determining what effect, if any, these trade-offs have, would likely require further perception studies of the sort we presented in section 3.5.

4.5.2 MSE Results and Discussion

The MSE of the estimators is plotted in figure 4.3. As in the bias plots, we can see that none of the proposed independent estimators have MSE performance similar to the best-case phase-corrected images. Efficient estimators should have MSE of 1 in these experiments, and we can see that best-case phase-corrected images represent the closest approximation to this of any estimator we considered.

As in our bias comparison, we can ask which estimator we would prefer if we excluded phase-correction. Again, this depends on what sort of trade-offs we are interested in making. The magnitude image (a in figure 4.3) has very high MSE at $s = 0$, but very quickly approximates the efficient estimator's MSE. Alternatively, the maximum stably adjusted profile likelihood (f in figure 4.3) is closer to the efficient MSE over more of the range, but does not approach the true efficient MSE as quickly.

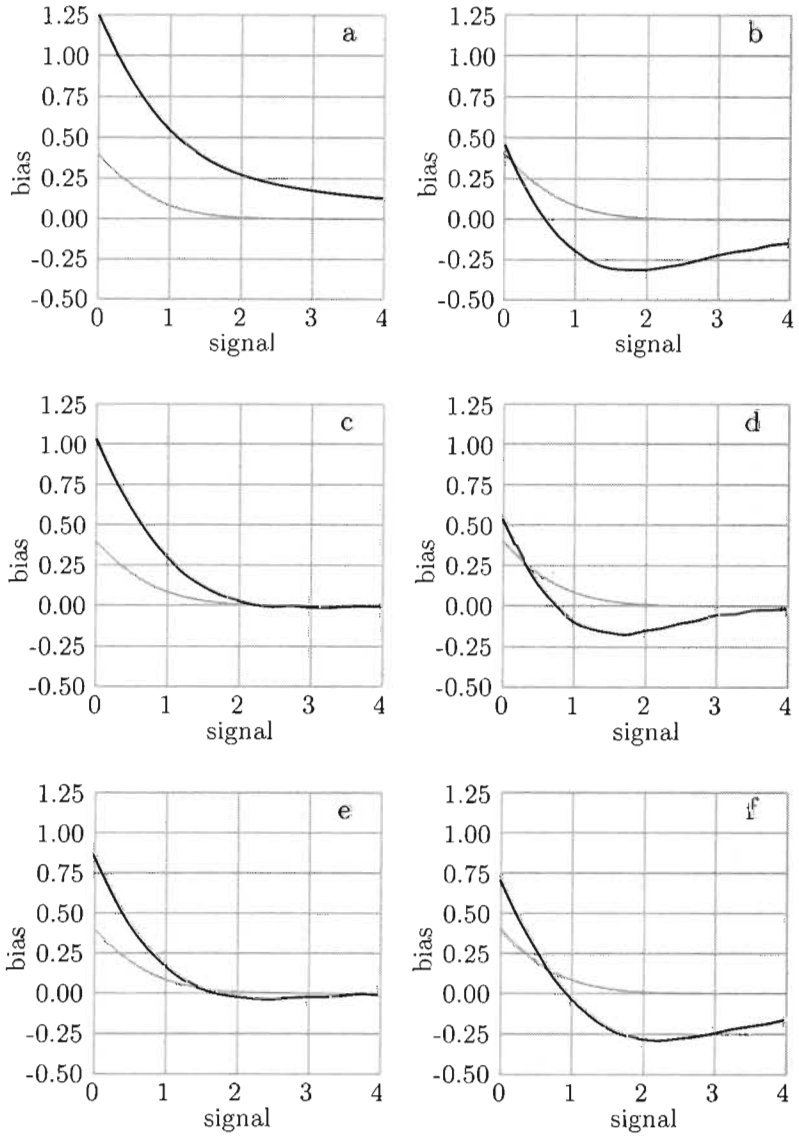

Figure 4.2: Biases of the estimators. Each plot displays the bias of one estimator (a: maximum profile likelihood, b: MM, c: GP, d: Sijbers/maximum marginal likelihood, e: maximum Bartlett-corrected profile likelihood, f: maximum stably adjusted profile likelihood) as a solid black curve. A lighter grey curve plots the bias of the best-case phase-corrected image estimator. An efficient estimator would have zero bias everywhere. The x-axis is the true signal value, and the y-axis is the bias of the estimator either computed directly or via the Monte Carlo experiments.

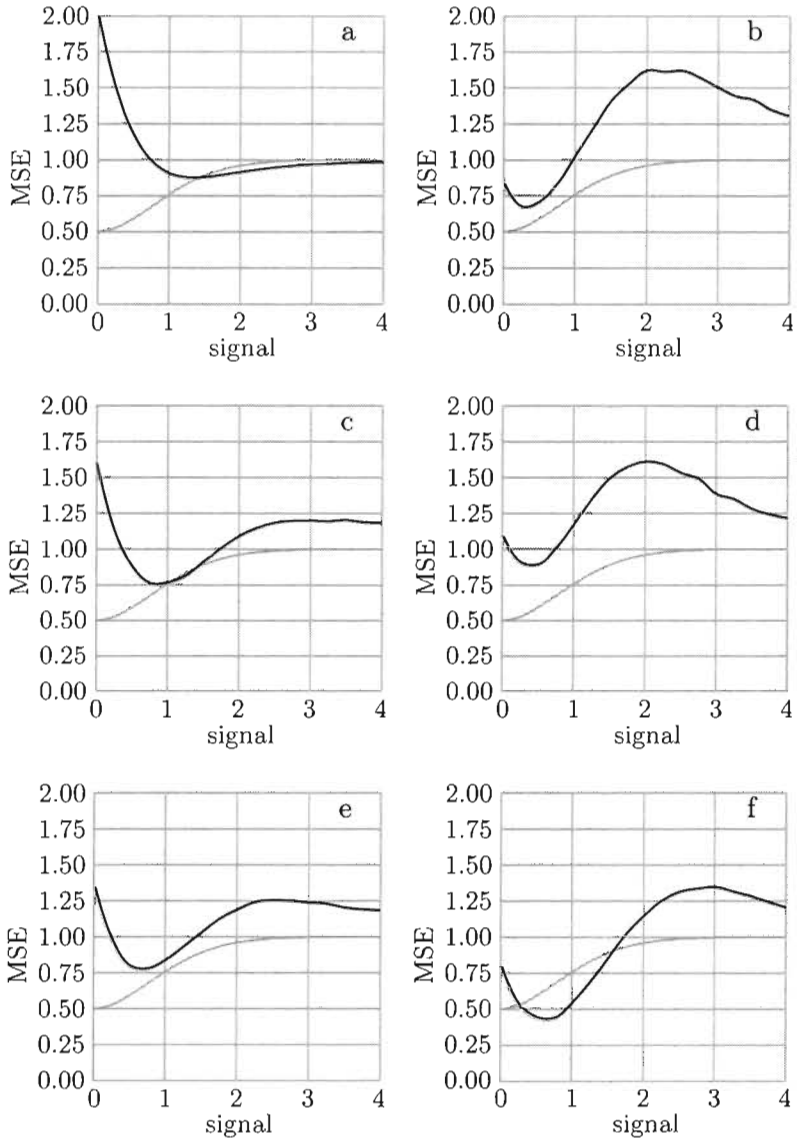

Figure 4.3: MSEs of the estimators. Each plot displays the MSE of one estimator (a: maximum profile likelihood, b: MM, c: GP, d: Sijbers/maximum marginal likelihood, e: maximum Bartlett-corrected profile likelihood, f: maximum stably adjusted profile likelihood) as a solid black curve. A lighter grey curve plots the MSE of the best-case phase-corrected image estimator. An efficient estimator would have MSE 1 everywhere. The x-axis is the true signal value, and the y-axis is the MSE of the estimator either computed directly or via the Monte Carlo experiments.

4.6 Conclusions About Independent Estimators for MRI

Comparing the results of the previous section we can see that none of the independent estimators derived via the statistical theory approximate the behaviour of the best-case phase-corrected real estimator. As such we cannot assume that our previous perceptual experiments can be used to justify the choice of any of these estimators. Moreover, the uniqueness of the phase-corrected estimator results imply that there is indeed a value in developing phase-corrected real estimators. We will pursue this further in chapter 5.

However, the results of these independent estimator experiments are also interesting on their own for several reasons. First, to the best of our knowledge there has not previously been a head-to-head comparison of the variety of estimators that are available for MRI. Additionally, we have found that the maximum Bartlett-corrected profile likelihood and the maximum stably adjusted profile likelihood estimators, which we are introducing to the MRI literature, are highly competitive with the estimators previously published. In addition to deriving new estimators, we have also related previously published estimators derived from the one-parameter model to the two-parameter model. Overall, our comparisons seem to indicate that, by choosing different independent estimators, we can trade-off bias for MSE at various signal levels. However, we are unable to draw strong conclusions from this comparison because it is unclear what trade-off is optimal at any signal level.

Based on our survey of the literature on independent estimators in MRI, we feel that our work highlights the necessity of being clear about which model is being used as the basis for estimation. For example, several authors have referred to their results as "the" maximum likelihood estimator for the MRI signal, but there are clearly several different maximum likelihood estimators, depending on the model used. This problem is accentuated when the estimation problem involves a more complicated model, such as the estimation of T_2 from a time series of decaying measurements, or fractional anisotropy from a set of directionally sensitive measurements. There are several papers in the literature where individual estimators are referred to as "the" maximum likelihood estimator for the T_2 estimation problem without making clear that they are producing the MLE for specific, simplified models [15, 70].

Our survey also seems to indicate that there is a belief that correct inference from the magnitude measurements r should be performed using the one-parameter model. This belief seems to be implicit in some work [34, 50, 55, 32], but is explicit in others [72, 71]. Our

understanding is that this model is used so that only the magnitude measurements remain and estimators will not rely upon the measurements of θ discarded by the scanner.

From our survey of the statistical literature, it is our belief that we are better off starting from the most complete model that we know is correct and then simplifying it using the theoretical tools for inference in models with nuisance parameters [36, 11, 8, 41]. In the case of MRI, the most complete model that we have at each pixel is the two-parameter model, and so we suggest this as the basis of estimation. Additionally, as our work shows, the resulting estimators do not rely on the measurements of θ and so can be used even after the data for θ has been discarded. This highlights an important and easily confused point in inference: measuring something and throwing it away is not the same as not knowing the measurement exists. In our case, our estimators derived from the two-parameter model know about the θ information and throw it away. The one-parameter estimators are derived from a frame of reference where the θ measurements are completely outside the model. Clarification on this issue in statistical inference would probably assist further work on other estimators in MRI.

Any of the estimators for which we have closed forms could be implemented trivially in an MRI data processing pipeline with minimal effort. Addition of the Sijbers estimator would require slightly more effort as its efficiency is highly dependent on the numerical root-finder employed at each voxel. In contrast to regularizing estimators of any sort, these independent estimators are very conservative and cannot introduce artifacts or smooth over fine features. As such, their use on medical images is likely to be low-risk. However, since we have not evaluated these estimators using perceptual experiments, it is not yet known whether they would enhance detection.

Finally, regardless of whether these estimators can enhance detection, we feel the introduction of statistical estimation techniques for models with nuisance parameters into the MRI estimation literature is useful. These tools could yield improved inference in quantitative estimation problems based on MRI measurements. For example, although the maximum profile likelihood [15] and maximum marginal likelihood [70] estimators for T_2 estimated from decay measurements has been presented, the full model for this problem has many nuisance parameters and might well be amenable to some of the techniques we have presented.

4.7 Summary

In this chapter we have introduced the basics of frequentist point estimation as a method of summarizing inferences from measurements about parameters. We have demonstrated techniques for dealing with multiple measurements and nuisance parameters when performing inference and defined the concept of an efficient estimator.

Based on our survey of the MRI literature, we have presented one- and two-parameter models that can be used for inference. We have surveyed the previous work using the one-parameter model and derived a selection of new estimators using the two-parameter model. In some cases inference from the two-parameter model assigns a new name to a previously published estimator, and in these cases we highlight how the previous estimator fits into the two-parameter theory.

We have compared all the estimators we presented using Monte Carlo experiments to measure bias and MSE. The estimators are evaluated not only relative to each other, but with the bias and MSE of the best-case phase-corrected real images we tested in chapter 3. We conclude that phase-corrected real images do have unique properties not replicated by any of the independent estimators. We also demonstrate that two of the new independent estimators we have derived are competitive with those currently published.

Finally, we analyze some of the reasoning behind the way statistical inference has been previously performed on MRI measurements. We highlight what we feel are limitations of these approaches and suggest some alternatives that may be broadly applicable in MRI estimators.

Chapter 5

Regularized Estimation of Signal Magnitude

Having considered approaches to independent estimation in chapter 4, we now focus on regularized estimators. As in the case of independent estimators, we can build regularized estimators using one- or two-parameter models. However, in the two-parameter case we now have an extra choice: regularization can be applied to either parameter individually, or both simultaneously. We will begin this chapter with a survey of estimators that regularize signal magnitude only. We will then proceed to discuss estimators that jointly regularize the phase and magnitude information. The structure of this work is very specific to the MRI signal and, interestingly, somewhat inconsistent with our approach to the two-parameter model. Having surveyed these alternatives, for the remainder of the chapter we focus on estimators that regularize only the phase parameter. We survey previous results on estimators of this type, and derive our own novel estimator. We demonstrate the result of our estimation approach using both imaging data and T_2 echo series.

5.1 Popular Regularization Methods

The fundamental idea of regularization is that we would like to impose some extra conditions on our estimates so that the values we infer for neighbouring locations are linked to each other. In the context of images, this is most commonly used to express the informal idea that neighbouring pixels are likely to have very similar values.

One of the earliest implementations of this idea imposes two cost terms that an estimate should simultaneously minimize. The first cost term penalizes an estimate more heavily the larger the 2-norm between the estimate and the measured data. The second cost term penalizes an estimate the larger the total squared gradient of the estimate. This combination of costs is known as Tikhonov regularization [4]. In general, this is not a good regularization scheme when the parameters being estimated contain edges and so it is not often used in image estimation. A more suitable alternative for image data is the total variation cost [60, 4]. This cost-function replaces the total squared curvature in the Tikhonov regularization formulation with the total one-norm of the estimate's gradient. In general this cost penalizes regions of slope and prefers homogenous regions with sharp edges at their boundaries. It has been suggested that this is a better model for edges.

Another method for enforcing regularity is to fit a polynomial to the data. Given real-valued measurements and a cost function that penalizes the 2-norm between the polynomial and the measurements, this can be fit efficiently using a weighted least squares algorithm. The difficulty in fitting polynomials is that changing any single coefficient will affect the estimate everywhere. An alternative is to fit local polynomials between the data points and insist on some degree of continuity between the polynomials. Thin plate splines have been presented as natural minimizers of cost functions that simultaneously enforce fidelity to image measurements (under specific definitions of fidelity) and minimize the total of the square of a chosen degree of derivative [86].

The linkage between neighbours can sometimes be better expressed by transforming the measurements to a different basis where the parameters are more compactly represented. Ideally, the parameters should only be represented using part of the basis, allowing us to infer that other elements represent pure noise. One simple approach to this is to represent the measurements in the Fourier domain, where natural images tend to be mostly represented in the low-frequency coefficients. Of course, by the properties of the Fourier transform we know that truncation in the Fourier domain is equivalent to convolution with some point spread function, and thus may not be good for images [16]. The wavelet basis has gained popularity as an alternative basis for representing measurements [22]. Unlike the Fourier basis functions, which have infinite support, wavelet basis functions have finite support but are still frequency-specific. Truncations in wavelet bases are, in general, not equivalent to simple convolutions, and so are less destructive to edges and better for image estimation.

Finally, a variety of methods use analogies to physical processes evolving towards a low-energy state. These methods begin with an initial candidate estimate, which is usually poor, and then use partial different equations (PDEs) to specify how the candidate estimate will change through an advancing pseudo-time. The pseudo-time is stopped when the evolution reaches a desirable candidate estimate, which is then taken as the estimated image. Due to their method of specification, these methods are referred to as *PDE-based*. Perhaps the simplest PDE we can imagine for this task is the heat-diffusion model. However, as with Tikhonov regularization and Fourier truncation, this PDE will tend to wipe out edges and so is not often desirable for image estimation. To address this shortcoming, anisotropic diffusion PDEs have been proposed [62, 4]. The classic example of this idea is to make the rate of diffusion inversely proportional to the gradient of the measurements at a point. This means that diffusion across edges will be slow, preserving them in estimates, while diffusion inside homogenous regions will behave like the heat PDE, smoothing the data.

5.2 Regularization Applied to the Magnitude Parameter

For researchers experienced in regularized image estimators, performing regularized estimation on the magnitude parameter of MRI is a natural extension of the work in other areas of imaging. Interestingly, only a small percentage of the published work applying regularization to magnitude images has focused on the one-parameter model described in section 4.2.2 [59, 64, 5]. As in the case of independent estimators, this model is chosen primarily from a desire to avoid estimators that might use the phase measurements discarded by the scanner.

In some cases, particularly in early applications of regularization to MRI data (e.g., [88, 28]), the one-parameter model is further simplified by assuming a Gaussian distribution for signal magnitude instead of the Rician distribution. The decision to use a Gaussian model stems from the existence of a wide variety of tools for smoothing Gaussian noise in images. One substantial difficulty in applying these techniques to MR images is that the Rician noise distribution is dependent on the true magnitude, where most models used for image regularization assume additive noise independent of the signal.

To address the Rician distribution, one common approach has been to work with the squared magnitude image [59, 64]. These estimators are thus the regularized cousins of the MM estimator discussed in section 4.3 — they first produce an unbiased estimate of s^2, and

then take the square root to produce an estimate of s. Although we addressed the point in chapter 4, since it is not clear in these publications it is important to reiterate that an unbiased estimate of s^2, regularized or not, does not imply an unbiased estimate of s by taking the square root.

5.3 Regularization Applied Jointly to the Magnitude and Phase Parameters

The application of regularization to both parameters has taken an interesting form in MRI. To the best of our knowledge no work has been done on regularizing s and ϕ directly. Instead, regularization has been applied to $s\cos(\phi)$ and $s\sin(\phi)$ by Wood and Johnson, Alexander *et al.*, and Bao and Zhang [91, 3, 6]. Consulting equation (4.8), we can see that changing the coordinates of the two-parameter model PDF's measurements back from polar to the real a and imaginary b coordinate system we have the two-parameter PDF

$$p(a,b;s,\phi) = \frac{1}{2\pi\sigma^2}\exp\left(-\frac{(a-s\cos(\phi))^2+(b-s\sin(\phi))^2}{2\sigma^2}\right). \tag{5.1}$$

Considering the structure of this PDF, the attractiveness of regularizing $s\cos(\phi)$ and $s\sin(\phi)$ becomes more apparent: if we create separate images of a and b, the real and imaginary measurements respectively, the result is two images whose true intensities are $s\cos(\phi)$ and $s\sin(\phi)$ respectively and each corrupted with independent zero-mean, additive Gaussian noise. As noted previously, there are far more regularization methods available for data with this noise model so a choice of well-developed techniques can be applied. Having estimated $s\cos(\phi)$ and $s\sin(\phi)$, Wood and Jonston, Alexander *et al.*, and Bao and Zhang propose to produce an estimated complex value at every voxel and then take the magnitude of this estimated complex value as the estimate of s [91, 3, 6].

While this approach is surely innovative, concerns have been raised about the methodology. First, neither of the parameters being regularized represent the information we are interested in. Instead, the parameter of interest and the nuisance parameter are mixed together and the result is regularized. Figure 5.1 shows the problem that results, where due to variation in ϕ, the majority of the signal alternately shifts between the real and imaginary channels of the spatial extent of the image. By splitting the signal power this way, we risk having our regularizing estimator perform more poorly in regions where $\sin(\phi)/\cos(\phi) \simeq \pm 1$.

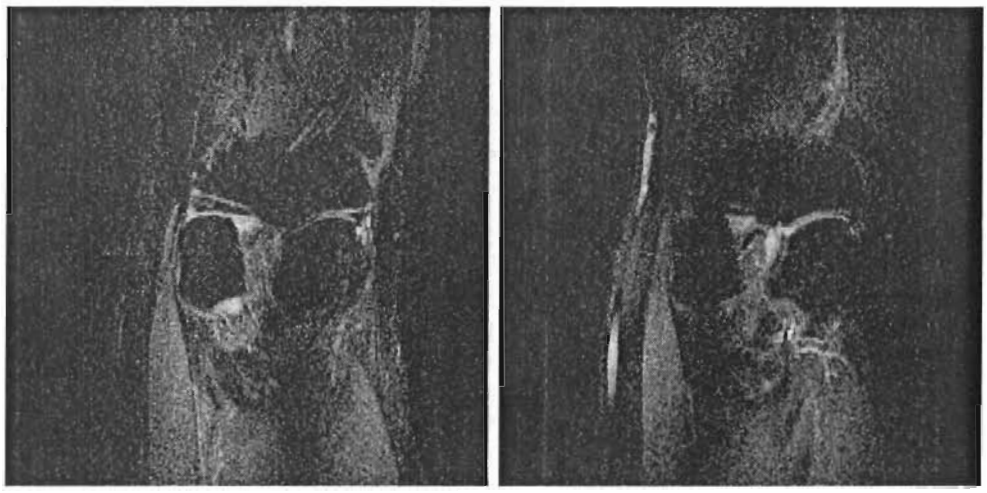

Figure 5.1: Absolute value of real and imaginary measurements of an inversion recovery knee image. *Left:* Absolute value of real measurements. *Right:* Absolute value of imaginary measurements. Due to the slow variation of ϕ across the spatial extent of the image, the signal alternately fades between the real and imaginary components. This is visible as black bands in the image where the signal is oriented completely in the opposite channel.

In these regions of the images, neither the real nor the imaginary image has a strong representation of the true signal and thus a greater amount of regularization may be applied to these measurements than is actually desirable. The desire to avoid this problem has been cited by other authors as one of the reasons they chose to developed regularized estimators from the one-parameter model [59].

5.4 Previous Estimators Regularizing the Phase Parameter

In chapter 2 we stated that the phase parameter varies smoothly across the extent of the image for a variety of pulse sequences. Due to this feature of the signal, the idea of regularizing the phase parameter has arisen in several different MRI applications. Although we are interested only in phase-corrected real reconstructions, the methods employed to regularize the phase parameter in other MRI problems are sometimes transferable to our problem. In this section we will first survey previous approaches to phase-corrected real reconstruction. In the second half of the section, we will present some of the phase regularization approaches that have been used in other MRI problems. In particular, we will consider the

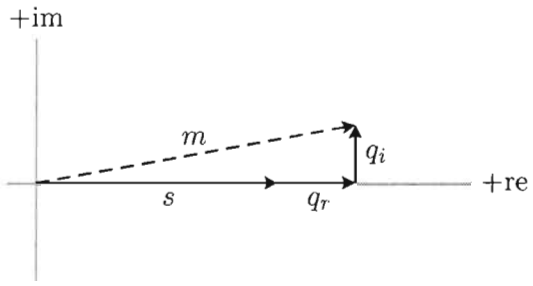

Figure 5.2: Example of a phase-corrected MRI voxel, showing the relationship of the various components in equation (5.2). The solid black lines represent the true parameter s and the noise in one sample from the phase-corrected distribution. The dashed line is the recorded value of the voxel while all the other parts of the diagram are unknown.

phase unwrapping problem and discuss its similarities to the phase correction problem.

5.4.1 Previous phase-corrected real estimators

Two-step phase-corrected real estimators

Previous publications on phase-corrected real estimators all share a basic formulation of the problem. In every case, the first step is to estimate ϕ with $\hat{\phi}$. The second step begins with the assumption that the estimate is the true value of ϕ. Based on this assumption, the measurements are multiplied by $\exp(-i\hat{\phi})$. As we can see in figure 5.2, if $\hat{\phi} = \phi$ this will rotate the true complex signal so that it is aligned with the real axis. If we call the real and imaginary components of the post-rotation measurement a' and b', and assume that $\hat{\phi} = \phi$, we can modify the two-parameter model shown in equation (5.1) to produce a one-parameter PDF

$$p(a', b'; s) = \frac{1}{2\pi\sigma^2} \exp\left(-\frac{(a' - s)^2 + b'^2}{2\sigma^2}\right) . \tag{5.2}$$

Figure 5.2 and this PDF both show that the post-rotation imaginary component b' contains no information about s and so can be discarded. Similarly, the post-rotation real component, a' is an unbiased estimator of s with a Gaussian distribution.

As in the case of independent estimation, multiple measurements at each location are assumed to be replaced by their complex average via an appeal to sufficiency. However, since none of the previous work we will consider relies on the PDF in the estimation of ϕ, we treat this data reduction more as a result of current scanner software (which will often perform

the complex averaging automatically) than as a necessary step in the phase-corrected real estimators.

All of the previous work shares identical approaches to data reduction and identical second steps. The only factor differentiating the algorithms is their formulation of the first step. We will now consider the two major families of previous approaches to estimating ϕ as step one of a two-step phase-corrected real reconstruction.

Polynomial-fitting

The first widely applicable approach to phase-corrected real reconstruction was presented by Ahn and Cho in 1987 [2]. In this work the authors assumed that the phase could be described by a polynomial of the form

$$\hat{\phi}(x) = c_1 + c_2 x\,, \tag{5.3}$$

where x represented the pixel index in the readout direction. The assumption here was that the principal components of ϕ were a constant offset, plus a linear factor proportional to the error in the timing of the signal readout. To determine the coefficient of this model, they proposed to use a weighted mean of the difference in phase measurements θ between neighbouring pixels as an estimate of c_2. Since Ahn and Cho were interested only in estimating phase variation in the readout direction, we can focus only on the rows of the image. Given N voxels in a row, we will refer to the measured complex value at the i^{th} voxel with magnitude r_i and phase θ_i. Ahn and Cho computed the weighted mean difference in a row via

$$\Delta = \arg\left\{\sum_{i=2}^{N} r_i r_{i-1} \exp\left[\mathrm{i}\left(\theta_i - \theta_{i-1}\right)\right]\right\}\,. \tag{5.4}$$

Averaging together the values of Δ for all the rows gives their estimate of c_2. Having corrected the linear component, they then suggested picking c_1 to center the histogram of partially corrected phases around 0. With a final estimate of $\hat{\phi}$ at every voxel via the polynomial, the authors then follow the standard second step in phase-correction and take $r\cos(\theta - \hat{\phi})$ as the estimate of s at each voxel.

An extension of this polynomial fitting approach to both x and y directions and higher-order terms, was made by Bernstein *et al.* in 1987 [14], who proposed a quadratic polynomial as sufficient for most practical data sets (the choice of quadratic polynomials has been suggested by other groups as well [46, 19]). However, this algorithm encounters a serious

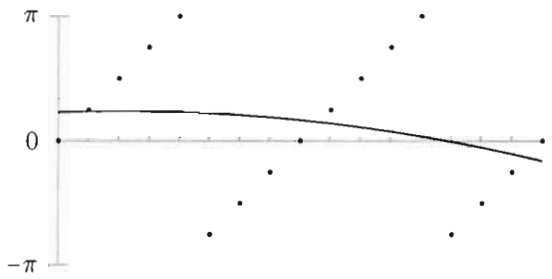

Figure 5.3: Demonstration of the fitting errors caused by treating angles as real values. Data values (shown as dots) are drawn from the function $f(x) = x$, sampled every $\pi/4$ from 0 to 4π and wrapped to the range $(-\pi, \pi)$. These are fit to a quadratic polynomial using linear least squares, resulting in the polynomial $0.72623 + 0.0340557x - 0.0105118x^2$. The error between the data and the polynomial demonstrates that even though the data function is linear, fitting a polynomial to the wrapped values is a non-linear optimization problem.

difficulty that afflicts most attempts to fit polynomials to phase measurements: the phase-data is circle valued. If the phase data were real-valued, the polynomial fitting would be solved by performing a linear least-squares fit. However, the difficulty with this approach lays in the computation of the distance between θ and the estimate $\hat{\phi}$. Their difference cannot be computed as $\theta - \hat{\phi}$ because they are angular values and thus points on a circle. Unlike on the real line, on the circle there are two geodesics (straight paths) between any pair of non-identical points. The error introduced by taking the real-valued difference is shown in figure 5.3. We can avoid this problem by taking the shortest of the two geodesics as our distance measure

$$\text{difference}(\theta, \hat{\phi}) = (\theta - \hat{\phi}) - 2\pi \left\lfloor \frac{(\theta - \hat{\phi}) + \pi}{2\pi} \right\rfloor , \tag{5.5}$$

this turns the polynomial fitting into a non-linear optimization problem. Bernstein *et al.* suggested fitting the constant and linear terms of the polynomial using the Ahn and Cho method and then using these as initial values for the non-linear search. Although there have been attempts to develop specific algorithms for this problem, a robust solution is not yet available [49].

An alternate approach to polynomial fitting has been present by Chang and Xiang [19]. Instead of using a non-linear optimization algorithm, they extend the differencing method proposed by Ahn and Cho to the estimation of coefficients for quadratic and higher terms. Chang and Xiang also introduce the novel idea of fitting the square of the complex image

in order to remove the sign from inversion recovery data at the cost of doubling all the phase errors. The major impediment to application of their algorithm is that it requires the quadratic and higher components to be small in the squared complex image [19]. However, when this is the case, their algorithm is quite efficient at estimating coefficients for low order polynomials.

Homodyne detection

The most common alternative to polynomial fitting is the *homodyne detection* approach presented by Noll *et al.* in 1991[58]. The core idea of this methodology is very similar to the joint regularization approaches discussed in section 5.3. The authors proposed to independently apply regularization to the real and imaginary measurements in order to produce an initial estimate of $s\cos(\phi)$ and $s\sin(\phi)$. This regularization normally takes the form of convolution with some kernel designed to remove the high-frequency information. However, unlike the approaches in 5.3, which take the jointly regularized data as the final estimate, Noll *et al.* use the phase of the regularized data only. This phase is taken as the estimate of the phase parameter ϕ and then, as before, $r\cos(\theta - \hat{\phi})$ is taken as the estimate of s. Criticism of this approach is very similar to that leveled against the joint regularization approaches, focusing on concerns about the mixing of the parameters [49, 51].

5.4.2 Other published uses of regularized phase estimators

The estimation of ϕ via regularized estimators has been used for a variety of applications in MRI. For example, phase estimation was used extensively in partial k-space imaging [46, 49, 58], although with the advent of multi-coil parallel imaging these methods are now somewhat less significant clinically. Additionally, the problem of phase unwrapping has received much work [43, 44, 20]. Many of the phase estimation algorithms applied to these problems can equally be refitted as part of a pipeline for the production of phase-corrected real images. We now survey proposed methods for these problems and focus on the differences between phase unwrapping and phase-corrected real estimation.

Partial k-space imaging

Partial k-space imaging [46, 49, 58, 44] was the principal means of accelerating MRI acquisition before the advent of multi-channel parallel imaging [74, 65, 31]. The basic idea

of partial k-space imaging is that, by taking advantage of the properties of the weighted proton imaging signal, we can avoid having to acquire up to half of k-space. Acquisition time is accordingly cut in proportion to the percentage of k-space that is skipped. We have already covered the first property of the imaging signal that is important to this idea: the information of interest is in the magnitude of the true signal and the true phase contains no information of value. This means that we could, in theory, represent our parameter of interest with just real values. Appealing to the properties of the discrete Fourier transform (DFT), we note that a real-valued signal will have a symmetric k-space [16]. This means that, if $\phi = 0$ everywhere, then we only need to measure one half of k-space and duplicate that half to produce the unmeasured entries.

The problem with this approach, as we have seen, is that $\phi \neq 0$ at all locations in almost any clinical image. However, if we know ϕ, we can multiply it out, as we did in step two of the phase-corrected real reconstruction, and then the phase-corrected k-space will be symmetric. Since we are more interested in the phase estimation than how the estimate is used in partial k-space imaging, we will not go into detail here about the various reconstruction approaches (for more details, see [49, 44]). Algorithms presented for this problem have generally fallen into the same two categories as phase-corrected real reconstructions: polynomial fitting and homodyne detection [46, 49, 51, 58]. In fact, most of the phase-estimation algorithms presented for this application were also presented as methods of performing phase-corrected real reconstruction.

Phase unwrapping

Phase unwrapping appears as a problem in MRI when, contrary to our description of weighted proton imaging, information is encoded in the true phase ϕ. For example, flow velocity can be encoded in ϕ using specific pulse sequences. However, since velocity is a real-valued quantity, when it is encoded in ϕ it may become wrapped if the range of encoded velocities exceeds the range $(-\pi, \pi)$. The core idea of phase unwrapping is to restore the real values from the angular data by taking advantage of the fact that ϕ should vary continuously in the image.

Many of the polynomial fitting approaches described for phase-corrected real imaging have also seen application in phase unwrapping [43]. The other common approach is to perform path-following through the image, attempting to determine the connectivity of each point to its neighbours (e.g., [20]).

There is an important distinction between phase unwrapping and phase-corrected real imaging. Phase unwrapping attempts to add integer multiples of 2π to images in order to form a continuous surface. Most often phase unwrapping is not concerned with estimating the true value of ϕ, but estimating it to within π so that the correct integer value can be added to the measurement [43]. By contrast, phase-corrected real imaging requires that we correctly estimate ϕ at each point. In this sense, phase-corrected real estimation is tasked with a greater challenge in terms of fidelity. However, the problem is also easier because our estimate of ϕ is indifferent to adding integer multiples of 2π. Thus, the two approaches are concerned with different aspects of ϕ. While it is possible to solve them both simultaneously, it is not necessary, which provides the intuition behind the derivation of our estimator.

5.5 Derivation of Our Novel Phase-Corrected Real Estimator

Having surveyed the previous work on phase estimation and phase-corrected real reconstruction, we can now present our novel algorithm for phase corrected real reconstruction. Our approach is motivated by an intuition about the problem that suggests a different approach; we begin this section by discussing these differences. Having clarified the reasoning behind our approach, we proceed to derive our novel estimator. However this formulation of the estimator has a free parameter that must be selected. We suggest a method for selecting this parameter from the data, producing a completely data-driven method for phase-corrected real reconstruction.

5.5.1 The intuition behind our approach

The basic intuition behind our problem is that almost all of the previous phase-corrected real reconstruction schemes unnecessarily perform phase-unwrapping. The chief exception to this is homodyne detection [58]. In general, applications of the regularization techniques covered in section 5.1 all perform unwrapping as an essential part of their operation. Our intuition is to suggest a regularization approach for phase estimation that does not simultaneously perform phase unwrapping.

Phase correction via polynomial fitting is the clearest example of the extra phase-unwrapping work implicit in current methods. Once we have fit a polynomial to the phase then we have implicitly unwrapped it. The polynomial is a map from voxel coordinates to reals which is then wrapped back onto the circle to get angular estimates of ϕ. Piecewise

polynomial fitting, which has received some study for application to angular data [47], has the same problem. Once we have defined a continuous surface as a function of the spatial coordinates, we have implicitly unwrapped the data.

Our intuition is that as long as an estimate of ϕ is constructed by restricting the derivatives of some continuous surface, estimation will require implicit phase unwrapping. This is because, by the act of constructing a continuous surface, the data is being unwrapped. Thus, even PDE-based methods designed for circular data [63, 77, 84, 18, 52] must implicitly solve the phase-unwrapping problem as part of estimating phase.

Based on this intuition, if we are to produce a regularized estimator that does not perform phase unwrapping, we must ensure that we do not implicitly construct a continuous surface. This means that we should not expect to fit one coherent model to the phase at all points in the image. The alternative approach we choose is to use, at each voxel, a local model with no guarantee about how this model relates to other voxels' local models in terms of continuity, smoothness, or similar properties. We have previously published a phase-corrected real estimator based on this idea [79]. However, in the remainder of this section we will present a new estimator that is based more rigorously on the point estimation theory we have presented in chapter 4.

5.5.2 The neighbourhood model

Having determined that we will construct our estimator using local models, we must define what local properties are to be captured and how regularity of ϕ will be expressed. Since the local model will be applied at each voxel, we begin specifying our model by defining the *voxel of interest* as having a neighbourhood of voxels denoted $\mathcal{L}$. Since we would like to perform inference inside each local model using the frequentist point estimation theory, we must specify a PDF for the measurements of the voxels in $\mathcal{L}$. From our description of independent estimation methods in chapter 4, we know that to perform inference without regularization, we would specify the PDF as the joint PDF of each voxel individually. If we were to use the two-parameter model (i.e., assume that σ has been estimated from the data and can be treated as constant), we would write this joint PDF as

$$p(\boldsymbol{r}_{\mathcal{L}}, \boldsymbol{\theta}_{\mathcal{L}}; \boldsymbol{s}_{\mathcal{L}}, \boldsymbol{\phi}_{\mathcal{L}}) = \prod_{i \in \mathcal{L}} p(r_i, \theta_i; s_i, \phi_i)\,, \tag{5.6}$$

where $\boldsymbol{r}_{\mathcal{L}}$ and $\boldsymbol{\theta}_{\mathcal{L}}$ are the collections of all measurements in the neighbourhood, $\boldsymbol{s}_{\mathcal{L}}$ and $\boldsymbol{\phi}_{\mathcal{L}}$ are collections of all the parameters in the neighbourhood, and the subscript notation on

the right of the equation is used to specify the measurements and parameters of individual voxels.

Although the local model relies on parameters from all voxels in the neighbourhood, it is being used to estimate only the parameters of the voxel of interest, which we will denote s_* and ϕ_*. Since the voxel of interest is one of the voxels in $\mathcal{L}$, when making inferences about s_* the joint PDF in equation (5.6) can be reduced by appealing to sufficiency. In particular, because of the independence of the measurements at each voxel, we can ignore all the other voxels in the neighbourhood and perform estimation using just the measurements in the voxel of interest. Clearly this is not what we want, since the whole purpose of applying a model to enforce regularity in ϕ_i at each voxel is to use information from the neighbouring pixels in the estimate of s_*. Thus, we must modify the joint PDF to express some sort of regularity.

Although many modifications expressing regularity are possible, we choose what is perhaps the simplest modification: force ϕ to be constant for all voxels in the neighbourhood, including the voxel of interest. The PDF that expresses this idea is

$$p(\boldsymbol{r}_{\mathcal{L}}, \boldsymbol{\theta}_{\mathcal{L}}; \boldsymbol{s}_{\mathcal{L}}, \phi) = \prod_{i \in \mathcal{L}} p(r_i, \theta_i; s_i, \phi)\,. \tag{5.7}$$

Note that there is only one ϕ parameter for the whole local model, and we have thus substantially reduced the degrees of freedom available to describe the measurements. With this formulation it is no longer possible to appeal to sufficiency and ignore any of the measurements in the neighbourhood. All of the measurements will help us perform inference at the voxel of interest.

Returning to the intuition behind the use of local models, we can now clearly see how this approach is different from those that define a continuous surface with local properties. Our local models cannot be joined to produce a coherent model for all the measurements unless we fix ϕ as constant across all voxels. Since we do not want to produce a coherent model, we assume only locally constant phase and perform estimation completely independently in each local model. Our local models do not influence each other except insofar as they overlap and use the same measurements for inference about different voxels of interest.

5.5.3 MLE of the neigbourhood model

Having specified the PDF of the local model that will be applied at each voxel of interest, we must now decide how to combine the PDF and the measurements to produce an estimate.

One obvious choice is the MLE of the magnitude of the voxel of interest in each local model. We can alternatively think of this as the maximum profile likelihood of the voxel of interest's magnitude. Since we know the maxima of these two functions are identical [8], in the derivations that follow we will solve for the MLE rather than construct the profile likelihood.

Locating critical points

Given the PDF in equation (5.7) we can invert the sense of the arguments and produce a likelihood function. Setting the score that results from this likelihood to zero gives the equations

$$\frac{\partial}{\partial \phi} \sum_{i \in \mathcal{L}} \ell(s_i, \phi; r_i, \theta_i) = 0 \tag{5.8}$$

$$\frac{\partial}{\partial s_i} \ell(s_i, \phi; r_i, \theta_i) = 0 \, i \in \mathcal{L} \, . \tag{5.9}$$

Thus, there is one equation of the first form and $|\mathcal{L}|$ equations of the second form, one for each of the $i \in \mathcal{L}$. Taking the derivative in each of the equations of the form (5.9) and rearranging gives us the MLE of the magnitudes under our local model

$$\widehat{s_i} = r_i \cos(\theta_i - \phi) \, . \tag{5.10}$$

We then perform the differentiation in (5.8) and substitute in the definition of $\widehat{s_i}$ given in (5.10) as the value of s_i in our further reduction to produce

$$\sum_{i \in \mathcal{L}} \cos(\theta_i - \hat{\phi}) \sin(\theta_i - \hat{\phi}) r_j^2 = 0 \, . \tag{5.11}$$

Using standard trig identities, this can be rewritten as

$$\sin(2\hat{\phi}) \sum_{i \in \mathcal{L}} \cos(2\theta_i) r_i^2 - \cos(2\hat{\phi}) \sum_{i \in \mathcal{L}} \sin(2\theta_i) r_i^2 = 0 \, . \tag{5.12}$$

Defining

$$u = \sum_{i \in \mathcal{L}} \cos(2\theta_i) r_i^2 \tag{5.13}$$

$$v = \sum_{i \in \mathcal{L}} \sin(2\theta_i) r_i^2 \, , \tag{5.14}$$

we can isolate $\hat{\phi}$ as

$$\hat{\phi} = \arctan(v/u)/2 + n\pi;\ n \in \mathbb{Z}\,. \tag{5.15}$$

Pairing this result with our estimate of $\widehat{s_i}$ shows that the choice of n does not change the location in the complex plane that is being estimated, only whether we locate it with a positive magnitude $\widehat{s_i}$ and a particular choice of $\hat{\phi}$ or a negative magnitude an the flipped choice of $\hat{\phi}$. These two descriptions of the point both represent maxima and thus valid MLEs.

Summary of MLE for local model

The results of our derivation show that the MLE of our local mode produces a two-step phase-correction approach. First, we estimate ϕ at every point using local models centered on each voxel. We then treat that phase estimate as if it was the true value of ϕ and, like the previously published phase corrected real estimators we presented, take $r_i \cos(\theta_i - \phi)$ as our estimate of s_i. Remembering that we are only interested in inference about s_* in each local model, we can write the MLE of the voxel of interest's magnitude, given the local model, as

$$\widehat{s_*} = \pm r_* \cos\left[\theta_* - \arctan(v/u)/2\right], \tag{5.16}$$

where the choice of $\pm$ is equivalent to choosing n even or odd in equation (??). This is similar to the pair of maxima that result from the MLE of the independent two-parameter model (see section 4.4.1). The MLE does not give us any guidance about which of these estimates to choose. We could solve this problem, as we did in the independent case, by forcing $\widehat{s_*} > 0$. However, we now have another alternative for choosing n by appealing to our assumption of regularity in ϕ, which can improve the quality of the resulting estimate.

However, before considering how to choose n, we would like to note one of the attractive features of equation (5.16) for implementation. We first note that equations (5.13) (5.14) can be rewritten as

$$u = \sum_{j \in \mathcal{N}_i} \left(a_j^2 - b_j^2\right) \tag{5.17}$$

$$v = 2 \sum_{j \in \mathcal{N}_i} a_j b_j\,, \tag{5.18}$$

allowing them to be calculated without trig operations. Further, with this definition calculating u and v for a local model surrounding every pixel becomes a convolution of an

indicator function for the neighbourhood with the two combinations of real and imaginary measurements shown in equations (5.17) and (5.18). Thus, having fixed the shape of the neighbourhood, the calculation of $\widehat{s_*}$, ignoring its sign, can be performed at all voxels in a number of operations that scales linearly with the data. This leaves the problems of estimating the sign at each voxel, and choosing the size of the neighbourhood.

5.5.4 Selecting between the two maxima

As we noted previously, the MLE does not give us a method for choosing between the two maxima of the likelihood function, and thus the positive or negative estimate of the voxel of interest's signal magnitude. While in the independent estimator we chose the positive estimate in every case, this is not necessarily desirable. It is impossible in weighted proton imaging for $s < 0$, but this does not mean that an estimator which produces a negative estimate of s is automatically incorrect. As we suggested when considering the independent MM estimator (section 4.3), a negative estimate of a non-negative is not necessarily wrong, but it is trivial to improve upon by simply truncating all negative values at zero. However, if we wish an unbiased estimator we might even ignore the truncation and accept negative estimates in exchange for being able to give unbiased estimates of $s = 0$.

Clearly it is desirable to get the correct sign for our estimate of s_*. Fortunately, given our assumption of ϕ as constant in local regions, there is a relatively simple method to choose the sign. We formulate the problem of selecting the orientation of ϕ_* at each point as an energy minimization problem. In our problem the states are n-odd and n-even and we can define our energy roughly, as encouraging a change of state at a point when doing so makes the phase closer to the phase of its neighbours in its local $\mathcal{L}$. More formally, we define the penalty assigned to a state of ϕ_* at a given point as the 2-norm of $s_* \exp(\mathrm{i}\phi_*)$ and the complex average of the neighbourhood (excluding the voxel of interest). If $s_* \exp(\mathrm{i}(\phi_* + \pi))$ has lower penalty, then we flip the orientation of ϕ_*. We simultaneously evaluate all voxels and flip them if necessary. We repeat this process until a stable state is reached and then take the resulting orientations at each point as our choice between the two available MLE maxima.

5.5.5 Choosing the neighbourhood

If we are to assume that our parameters can be approximated around any point with a region of constant true phase ϕ, it is natural to ask exactly how big these regions are and what shape they take. These questions are significant since the choice of $\mathcal{L}$ has a critical effect on the resulting estimate of $\hat{s}_*$ from each local model. If we exclude from $\mathcal{L}$ some pixels that really share a constant phase, we risk over-fitting the data and failing to take full advantage of the regularity inherent in our measurements. Alternately, if we choose $\mathcal{L}$ to include pixels that do not share an approximately constant phase, we risk under-fitting the data and mixing true signal with our estimated noise.

Leaving aside, for the moment, the question of the size of $\mathcal{L}$, we can begin by addressing the issue of its shape. Based on the previous experimental characterization of ϕ [46, 13, 49], we find that there is no *a priori* expectation of greater phase variability in one direction than any other. Thus, we should desire that $\mathcal{L}$ will be invariant under rotations, and so we chose it to be a circle centered around the voxel of interest. Pixels whose center is within the circle will be considered members of $\mathcal{L}$ and all others will be excluded. Voxels at the edge of the image will have have their neighbourhoods truncated as necessary.

Based on our choice of shape, we are left to pick the radius of $\mathcal{L}$. It is certainly possible that we could define a different radius to use in estimating ϕ_* in each local model. However, for the present work, we will choose one radius and use it to define all the neighbourhoods so that all the measurements can be used to inform our choice. To select the radius of our neighbourhoods, we will compute the error of our estimate at a variety of radii and then choose the radius with the lowest error.

We choose the cross-validation error of $s_* \exp(\mathrm{i}\phi_*)$ as our method of selecting the radius of $\mathcal{L}$. We calculate this error at each point by calculating the estimate specified in equations (??) and (5.16) and choosing the maxima that most closely matches the measured complex value at that point under the 2-norm. However, when we calculate u and v for each local model we do not include the voxel of interest in the neighbourhood. Thus, the difference between the measurement and the estimated complex value represents the prediction error using the local model to estimate the true parameters. Averaging this error over all of the voxels, each computed with its own local model missing the voxel of interest, gives us a complex value whose 2-norm we take as the cross-validation error.

5.5.6 Summary of complete algorithm

Having presented the individual components of our algorithms we will now summarize the order in which these elements are put together when presented with real MRI data.

Our first step is to separate the data into individual slices, since as we noted in section 2.4, in most clinical pulse sequences we cannot assume any regularity between slices. We then apply our algorithm to each slice individually. We begin the algorithm by determining the neighbourhood diameter that minimizes the cross-validation error. We then use neighbourhoods of this diameter to calculate the values of the parameters at each voxel of interest, using local models and equations (??) and (5.16). Having produced these initial estimates, we iteratively flip the phase estimates at each voxel until we have reached a stable energy state under our chosen energy function.

Having calculated our estimate of ϕ at every point, our algorithm now follows the same two-step structure as all other phase-corrected real reconstruction algorithms, taking the magnitude in the direction of ϕ as the estimate of s. If the final use of the estimated values of s at each voxel is the production of an image, we truncate the negative estimates of s at zero. If an unbiased estimator of s is desired, we take the aligned estimates of s as our solution without further processing.

5.6 Experiments with Real Data

In this section we present experiments with some different data sets that compare results using different phase-corrected real estimators with each other and the magnitude reconstruction. In particular, we will use the polynomial fitting algorithm approach suggested by Bernstein *et al.* [14], the homodyne detection approach suggested by Noll *et al.* [58] and the algorithm we are proposing. The Bernstein *et al.* algorithm has been implemented using the Levenberg Marquardt non-linear optimization algorithm in Mathematica 6.0 [90]. We have implemented the convolution in the homodyne detection algorithm by multiplying the DFT of the complex data with a function which takes the value 1 for a square of width w centered around the origin, and then takes values that linearly descend to zero over another $w/2$ DFT coefficients. Thus, all frequencies outside the lowest $2w \times 2w$ coefficients are set to zero. The smooth descent to zero is chosen to reduce Gibbs ringing. In each of our experiments we will specify what w was used. We will present experimental data sets that demonstrate the limitations of these approaches, and so we will not be particularly concerned with any

specific quality metric.

The goal of these experiments is not to demonstrate that our algorithm produces superior estimates to those already available. Instead, we are interested here in demonstrating the applicability of our general approach. Since our phase-corrected real estimator is the first built up from a local model employing frequentist point estimation theory, we are interested in demonstrating the feasibility of our approach.

5.6.1 Fluid-filled objects

The first data we consider is a 192×192 sample slice from a GRE volume acquired on a 3.0 Tesla Philips Achieva scanner. The three objects in the scene contained fluids (water and varying percentages of cream) which produce different values of s in this pulse sequence. The pulse sequence was selected so that signal from water and fat protons would be summed together in the true magnitude s positive sign. We used $w = 32$ for the homodyne detection algorithm.

The raw measurements and the results of the estimators are shown in figure 5.4. We can see from the results that both our estimator and homodyne detection have provided estimates that are relatively close to the original data. Of course, due to the low noise level in the measurements, we cannot see any improvement in image quality due to the phase-corrected estimation. However, we can see that the polynomial fitting algorithm of Bernstein *et al.* has failed to produce a viable estimate of ϕ, leading to incorrect estimation of s in the lower half of the image. We should note that this is not because it completely failed to fit the measured phases. In the top half of the image the fit is quite good. Thus, while perhaps there is a better quadratic fit to the phase measurements, it may also be the case that there is no quadratic that can accurately fit ϕ in this data.

To get a better understanding of how these estimators performed, we show the absolute differences between the magnitude image and our estimate, and the magnitude image and the homodyne detection estimate in figure 5.5. As we can see, the most significant error in both phase-corrected real reconstructions occurs around the edge of an air bubble at the top of the top-most water bottle. As we mentioned in section 2.3.2, air/water boundaries induce local variations in the magnetic field and thus local variations in ϕ. Since this region violates the assumption that ϕ varies smoothly, we expect that both of these methods will fail to produce correct estimates here. However, it is important to note that localized errors such as this do not induce errors in the rest of the image, and so we can see that both of these

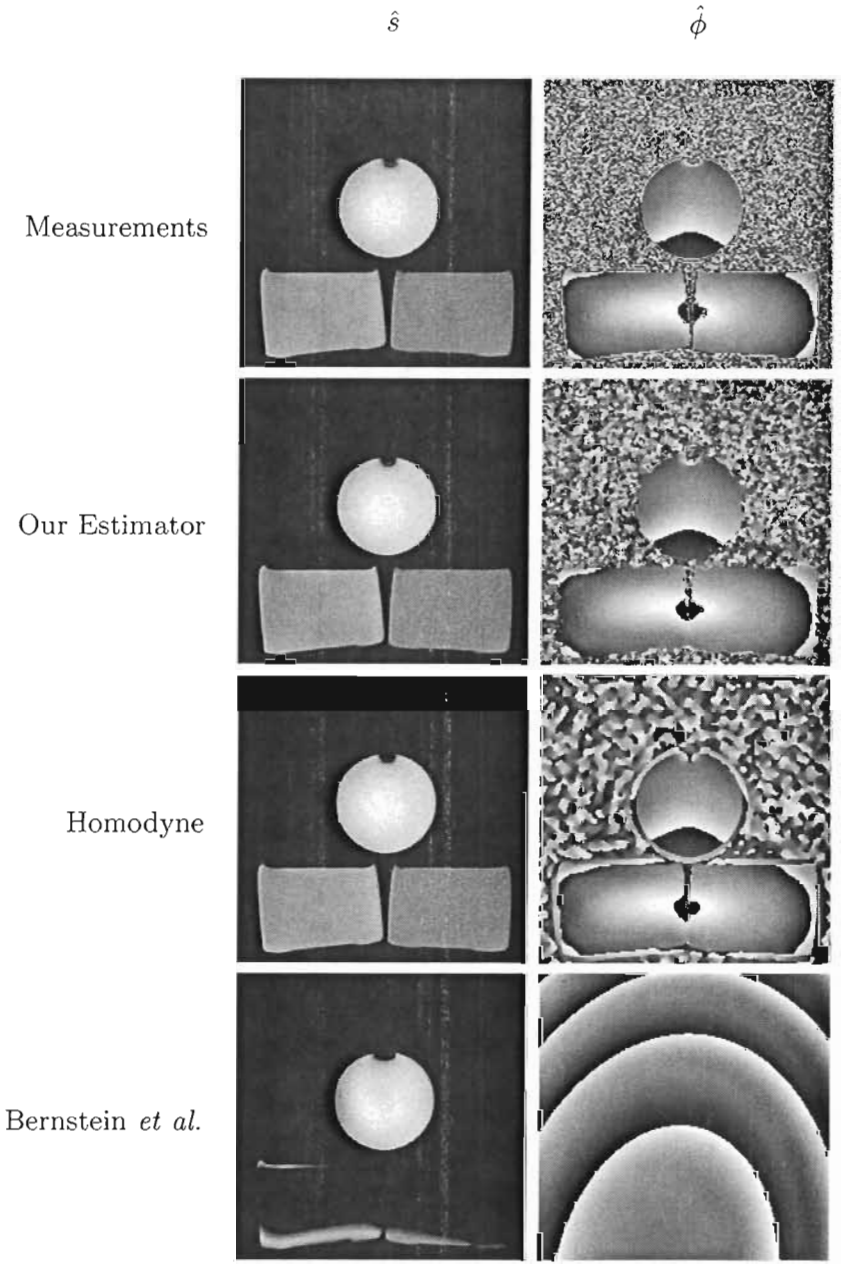

Figure 5.4: Magnitude (left) and phase (right) images of the slice chosen from the fluid-filled object volume for our experiment. The first row shows the measured values which are the MLE of s and ϕ. The remaining rows show the positive part of the estimate of s and the associated estimate of ϕ produced using the three phase correction algorithms.

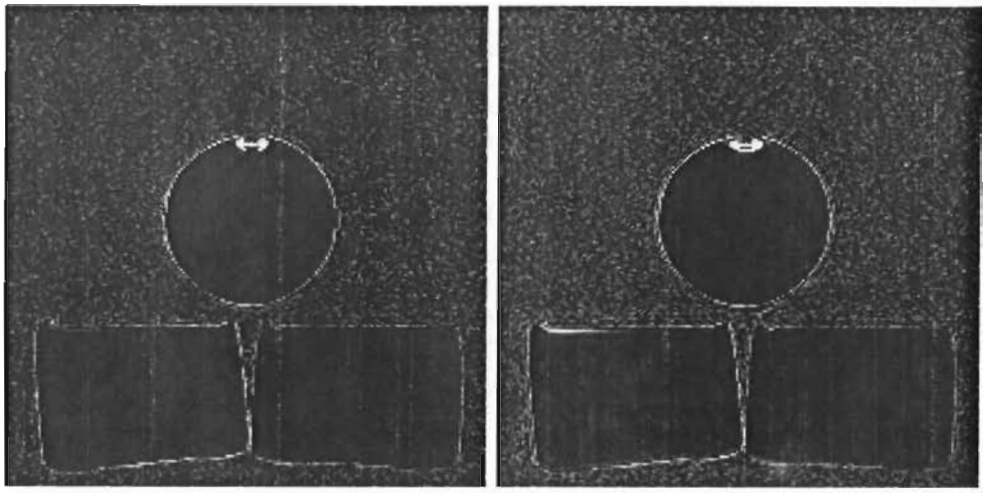

Figure 5.5: Absolute difference between magnitude and phase-corrected estimates of fluid-filled object slice. The difference between the magnitude and our estimate is shown on the left. The difference between the magnitude and the homodyne detection estimate is shown on the right. Both images are in the same units for accurate comparison, and contrast has been reduced to compress small and large errors into the same image. Note that the principal difference that occurs in both images is the error around the water bubble in the top-most object.

methods are relatively robust when presented with local violations of their assumptions.

5.6.2 Fat-suppressed knee

Our second experiment uses a fat-suppressed knee image. This data has a slowly varying value of ϕ and so is not particularly challenging for any of the available estimators. We include this data because it has an interesting artifact that demonstrates differences between our approach and the other two we present. Fat is suppressed in these images using an approach called short-tau inversion recovery (STIR) in which the pulse sequence should ensure fat makes no contribution to the value of s at any point. However, due to magnetic field inhomogeneities, there are a few regions of fat that appear in the image. These fat regions form an interesting artifact because the signal is inverted and thus the sign of s is negative. We are interested in how the available approaches estimate the signal in the fatty regions.

In figure 5.6 we show the raw measurements and the estimates of the positive and

inverted components of s, along with the estimated phase. To make clear where the regions of inverted fat exist, we have put two arrows on the magnitude image. As we expect, all three of the phase-corrected real estimators assign a negative estimate to the fat signal at these points and thus their estimated images are black in these regions. However, understanding how each estimator arrived at this estimate for the fatty regions illustrates their differences. For example, the polynomial fitting algorithm proposed by Bernstein *et al.* does not include any model of inverted signal. Thus, when the patches of inverted signal are small, like in this image, they are simply considered error by the fit. If the patches of inverted signal were larger, they would likely begin to affect the estimated value of ϕ. Our algorithm, on the other hand, first estimates the fat as being positive signal, and then orients it negatively during our iterative orientation selection stage.

In figure 5.7 we show the raw measurements and the estimates of the positive and inverted components of s, along with the estimated phase produced by the homodyne detection estimator at two different values of w. At $w = 2$ the homodyne detection algorithm behaves like the polynomial fitting algorithm, smoothing over the small regions of inverted signal in favour of the larger regions of positive signal. However, at $w = 16$ the fitting is now fine-grained enough to fit the patches of inverted fat signal as separate regions, estimating them with positive signal instead. We show this result to demonstrate the importance of choosing a correct window size when using the homodyne detection algorithm. In our previous data (figure 5.4) $w = 32$ was a correct setting, while here we find $w = 2$ to be correct. Further, using $w = 16$ on this knee data results in the erroneous orientation of the fat signal illustrated by the arrows. Contrary to this, both the polynomial fitting approach and our algorithm do not require the choice of a parameter like w. Of course, this is not to say that there is no way to automate the choice of w in homodyne detection. However, to the best of our knowledge no such scheme has been presented in the literature.

5.6.3 Discussion of experimental results

We have now seen two example data sets that highlight the differences between our estimator and the approaches previously presented in the literature. It is not difficult to construct pathological synthetic data that will cause any of these algorithms to fail, so our goal here is not to choose one algorithm as better than the others. Instead, we believe these experimental results demonstrate that the use of local models and our novel, statistics-based approach to phase-corrected real reconstruction is a viable alternative to polynomial

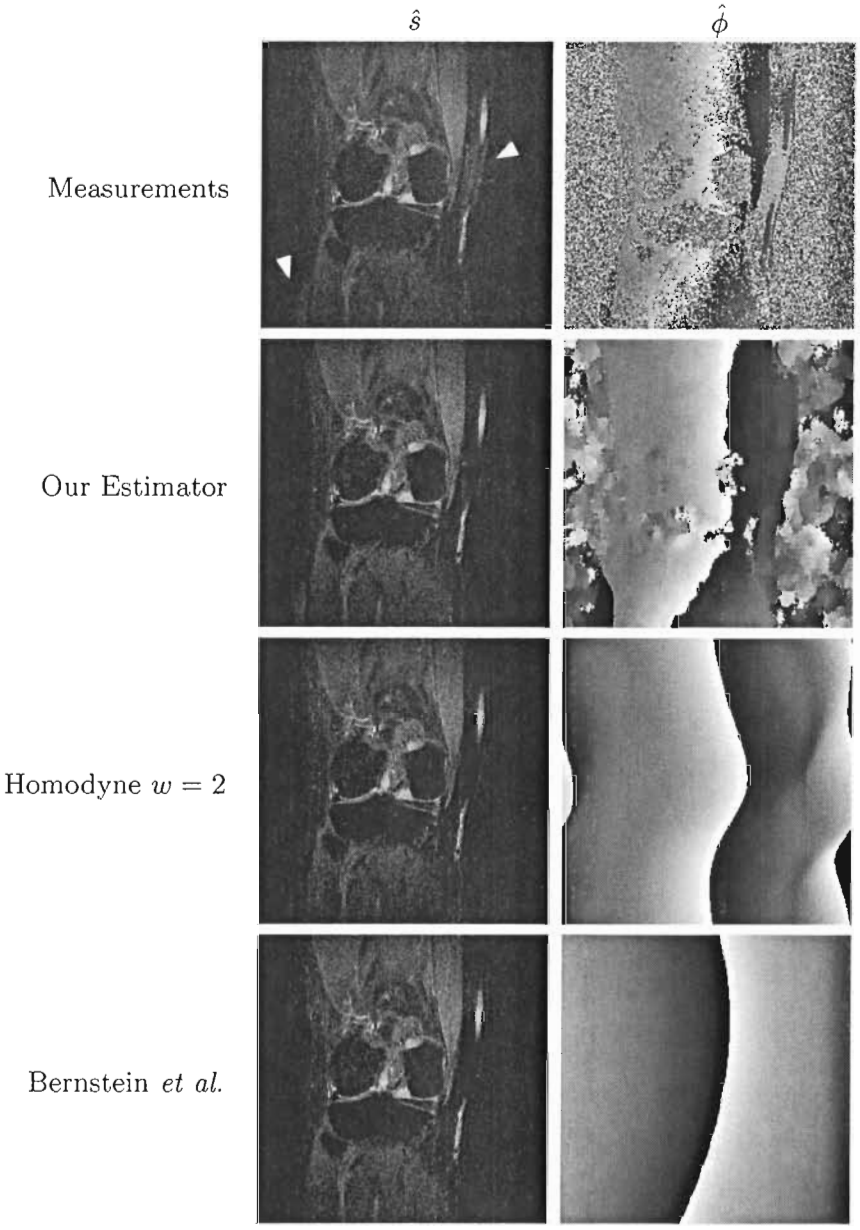

Figure 5.6: Magnitude (left) and phase (right) images of the slice chosen from the knee volume for our experiment. The first row shows the measured values which are the MLE of s and ϕ. Arrows on the magnitude image indicate the locations of inverted fat signals. The remaining rows show the positive part of the estimate of s and the associated estimate of ϕ produced using the three phase correction algorithms.

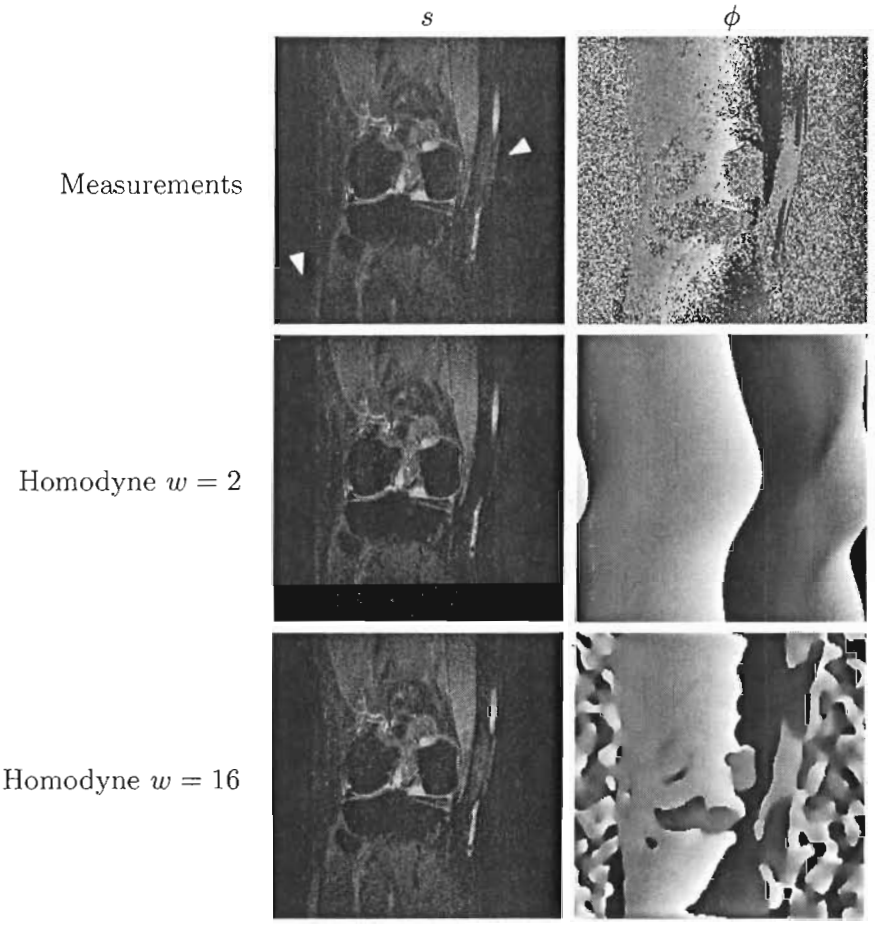

Figure 5.7: Magnitude (left) and phase (right) images of the slice chosen from the knee volume for our experiment. The first row shows the measured values which are the MLE of s and ϕ. Arrows on the magnitude image indicate the locations of inverted fat signals. The remaining rows show the positive part of the estimate of s and the associated estimate of ϕ produced using the homodyne detection algorithm with different settings of w. Note that changing the value of w alters whether or not the fat is included in the estimated positive image.

fitting and homodyne detection. Our algorithm has demonstrated general correct behaviour with local failure when confronted with data violating its assumptions, a core property of the use of local models and one that we feel is important in analyzing clinical data. Additionally, our local model approach allows the algorithm to locally determine the orientation of each voxel, making it feasible for our estimator to orient fat and water properly in our example data in a second step after an initial estimate of ϕ is constructed.

5.7 Application to T_2 Measurements

The measurement of T_2 echo series suggests an interesting application of our approach to phase-corrected real reconstruction. As discussed in section 2.5, the bias introduced by Rician noise can be quite substantial in later echoes when there is little signal remaining. Given that many of the analysis algorithms used to process these echo series assume Gaussian instead of Rician noise [89, 15, 76], it seems reasonable that phase-corrected real estimation can bring the data into line with their noise assumptions. Thus, while in the imaging case we truncated all our estimates of s below at zero, in this problem we will want to keep all the negative estimates of s to ensure that we have an unbiased estimator at each echo.

In this section we demonstrate a phase-corrected real estimator for 32-echo T_2 series where the values of t are equispaced between echoes. Before proceeding to apply phase-corrected real reconstruction to our T_2 data, we need to produce a model of our measurements. We could expect that the magnitude parameters in a given echo series are related by equation (2.5) with only our chosen variation in t between echoes' magnitudes. However, we will ignore this relationship between echo magnitudes in developing local model for phase-corrected estimation and treat all the echoes' magnitude parameters as completely unrelated. We do this so as not to bias any later T_2 analysis algorithms for which our estimates will be used as input.

However, we would like to assume regularity between echoes in the parameter ϕ. Our discussions with physicists suggested that a given voxel should have two values of the phase parameter ϕ; one for the even-numbered echoes and one for the odd-numbered echoes. While it was true that in our measurements the phase of the odd and even echoes were generated with different processes, we found that our measurement indicated that ϕ was a linear function of the echo number, as shown in figure 5.8. Having confirmed our hunch with our physicist collaborators, we used this linear model of phase to derive an experimental

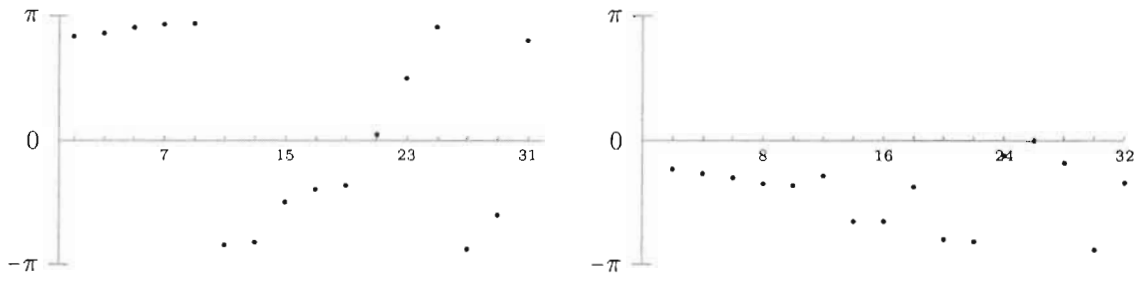

Figure 5.8: Phases of odd (left) and even (right) echo series in water sample. Phase is plotted on the y-axis and echo number advances from left to right in each plot. The phase of the early echoes is clearly a linear function of the echo number, with constants such that the phase is increasing with echo number in the odd echoes and decreasing in the even echoes. This pattern is eventually overwhelmed by noise in the later echoes. However, the deviations in later echoes do not mean that the underlying true phase ϕ is not still linear, just that the measured phase θ is now more affected by noise than by the true phase ϕ.

algorithm for T_2 fitting. Note that this algorithm is likely highly specific to the combination of pulse sequence and scanner used to measure this data. It is for this reason that we highlight the experimental nature of this algorithm and we present it as an example of how our local models and statistical methods can be extended to other problems in MRI.

In the following application we will demonstrate our algorithm on one voxel in parallel with our explanation of the algorithm. This voxel comes from a bottle of water and is known to have $T_2 = 25$ ms. Its phase measurements are shown in figure 5.8 and its magnitude measurements are shown in figure 5.9.

5.7.1 Phase-corrected real estimation of T_2 data

We begin with 32 echoes, and split them into odd and even echoes which will be treated separately since their ϕ parameters come from different processes. Given that we have 16 echoes in the odd and even groups at each voxel in our data, we do not need to use a local model that imposes any sort of regularity between neighbouring voxels. Instead we will impose a local model on the odd and even echoes at each voxel that assumes the ϕ parameter varies linearly with echo number. We begin by estimating the slope of ϕ with respect to echo number for the odd and even echoes using the method presented by Ahn and Cho [2]. Following their example, we then correct for this linear shift across the echoes at each voxel. The odd and even phase data of our experimental voxel is shown after the

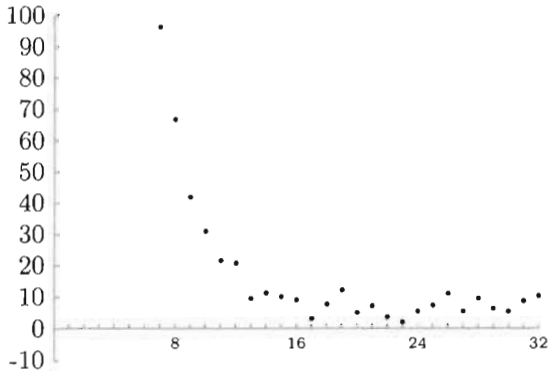

Figure 5.9: Magnitude of echo series in water sample. Magnitude is plotted on the y-axis and echo number advances from left to right. Early echoes have magnitudes greater than 100 and so are not plotted to preserve scale. Note the decay due to T_2 weighting and the positive bias in the later echoes due to Rician noise.

linear correction in figure 5.10.

As the example in figure 5.10 shows, once the linear correction is performed we are relatively confident that one ϕ is fixed at any given voxel for all of its odd echoes and similarly for the even echoes. We can then apply our local model of constant phase that we outlined in the last section to the odd and even echoes independently and compute the MLE of all the elements in the odd and even echoes series of the voxel. Given that the local model applied to all odd or even echos at a voxel is the same, we can perform simultaneous estimation treating the magnitude of all odd or even echoes in the voxel as the parameter of interest. We note that this is different from our previous use of local models to enforce spatial regularity in which the local model could only be used to estimate the voxel of interest.

Having independently estimated the parameters of the odd and even echoes in a voxel, we interleave the estimates to produce an estimate of the complete echo series. However, we note that the MLE for this model still does not allow us to chose the orientation of our echoes automatically. We address this problem by aligning our odd-echo estimates at each voxel to minimize its 2-norm with the complex measurement of the first echo and our even-echo estimates to minimize the 2-norm with the second echo. We choose the first echo and second echoes as our targets for alignment because they are guaranteed to be the odd and even echos with the highest value of s and should be completely uncorrelated with noise in later echoes. The phase of our sample voxel after this orientation correction can be seen

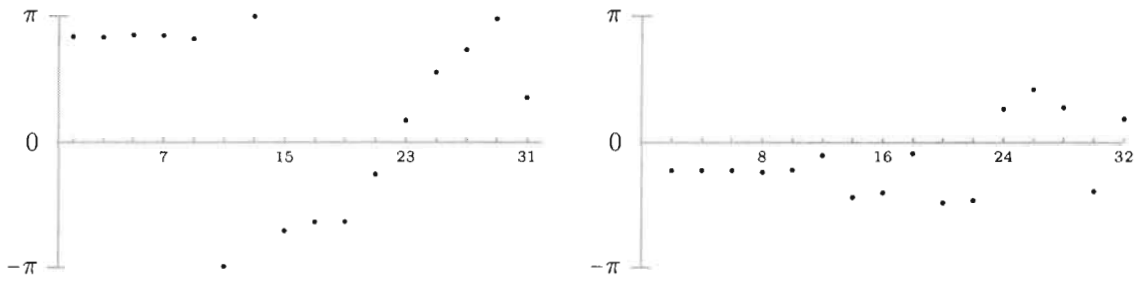

Figure 5.10: Phases of odd (left) and even (right) echo series in water sample after linear phase correction. Phase is plotted on the y-axis and echo number advances from left to right in each plot. The early echoes now clearly show the underlying constant value of ϕ. As in figure 5.8, later echoes are corrupted with substantial amounts of noise and so the constant value is obscured.

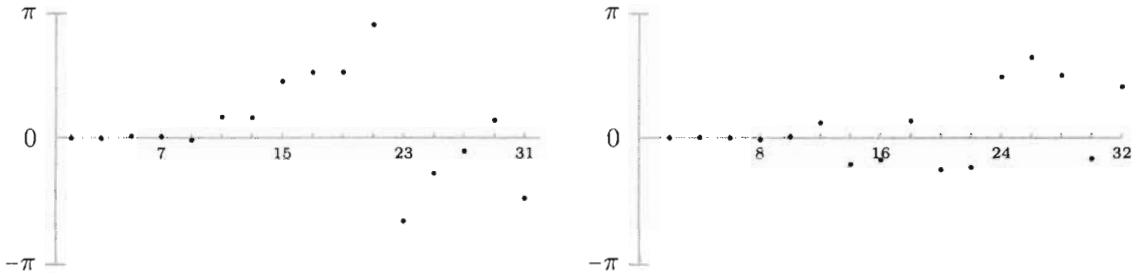

Figure 5.11: Phases of odd (left) and even (right) echo series in water sample after complete phase correction. Phase is plotted on the y-axis and echo number advances from left to right in each plot. The phases of the early echoes are now approximately zero, with deviation from zero increasing due to increased noise in later echoes.

in figure 5.11. Our phase-corrected estimate of s at each echo is shown in figure 5.12. We can see in figure 5.12 that our estimates are now far closer to unbiased than the magnitude echo series, even as the true value of s becomes zero in the later echoes (see later echoes of figure 5.9 for comparison).

5.7.2 Discussion of results

The ability of phase-corrected real reconstruction to produce unbiased echo trains in T_2 data is depicted in figure 5.12. Although we will not plot other echo series here, it is the case that all other exploratory plots we made from the estimated echo series in voxels with short relaxation times demonstrated the same unbiased property in later echoes. This result is

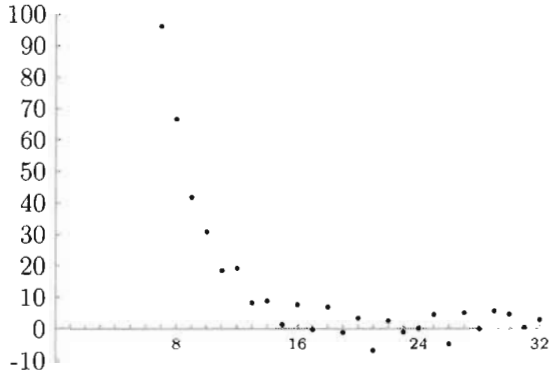

Figure 5.12: Phase-corrected estimates of s at each echo in water sample. Estimate of s is plotted on the y-axis and echo number advances from left to right. Early echoes have phase-corrected estimates greater than 100 and so are not plotted to preserve scale. Note that the later echoes now have mean zero and so are unbiased.

significant in that several papers on T_2 curve fitting have focused on dealing with bias during data analysis [70], or justifying why the bias can be ignored [15]. By providing unbiased, Gaussian-distributed estimates of the echo magnitudes we can avoid the necessity of either of the previous approaches.

Interestingly, further inspection of the data highlights some properties that suggest further improvements to our model are possible. The differences between the phase of the even and odd echoes, hypothesized to be completely independent, looks to be quite close to π in figure 5.10. Additionally, there appears to be some correlation between the even and odd neighbouring echoes in figure 5.12. These correlations indicate that there is likely some underlying parameter driving both the even and odd echoes phases, and that specification of this model can further improve the results using our general approach.

Further work on applications of this data to problems using T_2 measurements will be necessary in order to meaningfully quantify any improvement that comes from this type of method. Additionally, we have used a very simple model to perform these estimates. A more complicated estimator could factor the T_2 parameter explicitly into the model or take advantage of similarities in ϕ between neighbouring voxels and not just within echo series. However, despite the preliminary nature of these results in terms of their use in T_2 based problems, we view these estimates as further evidence that our approach to phase-corrected real estimation can produce improvements in MRI data.

5.8 Summary

We began this chapter with a discussion of the various forms of regularization that are commonly applied in image estimation. We then surveyed the various ways in which regularizing estimators have arisen in the context of MRI. We focused particularly on the problems in which estimation of the phase parameter ϕ is important. Having considered the previous work, we described our intuition about how to separate the phase-corrected real estimation problem from the phase unwrapping problem. We suggest that fitting global models to the data forces us to solve the phase unwrapping problem in parallel with the phase-corrected real estimation problem. Since we are only interested in the latter, we should derive an estimator that eschews global modeling.

Following this intuition, we used the point estimation theory and models we presented in chapter 4 to derive a novel phase-corrected real image estimator using independent local models at every voxel. This is followed with some experiments demonstrating the feasibility of the algorithm we have derived compared to existing algorithms. Finally, we conclude the chapter by applying the same general estimation methodology to T_2 echo series and demonstrate the resulting unbiased estimates of the signal parameters. These experiments suggest that promising phase-corrected real estimators can be derived using local models and point estimation.

Chapter 6

Conclusions

6.1 Summary of Contributions

The research program explored in this thesis has produced three major contributions. First, we have demonstrated a method for using high-field MRI data and a model of the MRI signal to evaluate MR image estimators using 2-AFC experiments. Second, we have rephrased the independent MRI estimation problem in terms of the statistical literature on point estimation in models with nuisance parameters, highlighting the rich theory on this problem produced by statisticians. Third, we have used point estimation theory and local models at each voxel to produce a novel phase-corrected real estimator which avoids the phase unwrapping problem.

Although there has been previous work using task-based quality measures to evaluate image quality in other medical modalities, to the best of our knowledge our work is the first application of this approach to the evaluation of MR images. Previous approaches have focused on 2-norm error measures or task-free perception-based quality measures. However, none of the previous work has presented a method to produce synthetic MRI images for experiments, making it difficult for researchers developing estimators to evaluate these algorithms. Our work has combined our understanding of the MRI signal's characteristics with the previous work on image quality in other modalities to demonstrate a viable method for evaluating MRI estimators using task-based experiments. Using this work, we have provided empirical corroboration for the previous theoretical assertions that phase-corrected real images could improve viewers' ability to detect signal in dim regions of MR images.

We have shown that, despite the range of publications deriving independent estimators

for the true magnitude parameter s, very few of these approaches are based on the methods of inference developed by statisticians. Additionally, in the cases where the MLE has previously been applied to MRI data, it has almost always been applied to a reduced model that does not fully describe the true properties of the signal. We have demonstrated that the estimation of s from MRI measurements can be addressed using point estimation theory applied to the full model of the MRI signal. Using these methods we have derived two novel estimators and related several existing estimators to the point estimation theory. In doing this we have demonstrated that the model reductions employed by previous authors are both unnecessary and, further, incorrect under the logic of frequentist statistical inference.

However, none of the independent estimators that we considered performed like a phase-corrected real reconstruction. Thus, we explored application of the point estimation theory in such a way as to enforce local regularity but not perform phase unwrapping. We proposed the application of independently applying regularity-enforcing local models at each voxel and then performing point estimation using the MLE. The result of this method is a novel form of the two-step phase-corrected real reconstruction algorithm formulation seen in several previously published variations. Applying our result to empirical data, we find that our method is competitive with several of the alternative approaches published previously. Additionally, we show how the basic intuition of our method can be applied to the estimation of T_2 echo series. Considering these two results together, we suggest that our introduction of point-estimation theory and independent application of regularity-enforcing local models as a method for deriving phase-corrected real estimators is promising. This method suggests a general framework that can be expanded upon to derive phase-corrected real estimators for a wider variety of MRI estimation problems.

6.2 Future Work

Much of the work in this thesis has involved relating different areas of research to suggest new methods for problems in MR data analysis. The approaches that we have proposed suggest a variety of possible avenues for future research. We outline some of these ideas below, grouping them into the major components of our research program.

6.2.1 Task-based image estimator quality experiments

While the results of our task-based estimator evaluation experiments convinced us that proceeding with phase-corrected real estimation was a useful course of action, there are several factors in the experiments that could be improved upon in future studies. Perhaps most importantly, we would like to improve the experimental interface so that users can adjust image contrast and brightness in realtime the same way radiologists do when viewing images. The adjustments to contrast and brightness essentially allow the users to explore a limited range of image estimators based on the estimator applied to the original measurements. This exploration could affect how successful users are at locating objects in images and so we feel that any future experiments should include this feature.

In broader terms, the widespread adoption of these sorts of experiments by researchers investigating estimators requires that software tools be made available to reduce the large amount of labour required. In particular, we feel that a tool to assist investigators in producing test images using their own experimental estimators would be of great use. There are several groups currently working on providing experimental workstations for displaying images and collecting data as we have done in our experiment, but as of yet there is no easy way for a researcher to easily generate sufficient examples to run these experiments. Further, standard implementations of the model observers would likely encourage more researchers to publish results comparing them with human performance, and thus enhance our understanding of how these models can be used to evaluate image quality.

6.2.2 Independent estimators

Although we found that none of our estimators approximated the performance of the phase-corrected real estimator, this does not necessarily mean that they are not an improvement over using magnitude images. We feel that the maximum Bartlett-corrected profile likelihood and the maximum adjusted profile likelihood estimates are worth evaluating using task-based experiments due to their ease of practical implementation and their substantially different behaviours compared to the magnitude estimator currently used.

Of course, the estimators we have derived are by no means the exhaustion of all the possible statistical theories for dealing with nuisance parameters. Although we have presented what we feel are in most common approaches to the problem, there are likely many more estimators that can be derived by applying tools from the statistics literature to our

problem. Additionally, we have not explored whether Bayesian reasoning about the problem would suggest different approaches, beyond noting that the MLE and maximum marginal likelihood estimator both have equivalent formulations in Bayesian inference.

Finally, a very interesting problem that arises from estimation is how to display inference producing confidence intervals at each voxel. In almost all frequentist estimation problems, confidence intervals are preferred to point estimates as methods of summarizing the result of inference. While confidence intervals can be calculated for our parameter of interest at each voxel, it is not clear how to display these intervals in a meaningful way to radiologists. Improving the visualization of the uncertainty in estimated medical images seems like a potentially fruitful problem, particularly in modalities like low-field MRI where there is a substantial amount of uncertainty in the estimate.

6.2.3 Phase-corrected real estimation

There are several natural routes for further development of the phase-corrected real estimator we have developed. The two that are likely to have the largest improvement on the quality of the output are locally linear phase models and improved orientation correction.

Modifying the local model to assume locally linear phase variation instead of locally constant would likely provide significant quality improvements as the size of the neighbourhoods used for inference could be substantially expanded. Noting that a linear phase shift in the spatial domain is equivalent to a coordinate shift in the Fourier domain, and that a real-valued signal should have a symmetric representation in the Fourier domain, there may well be efficient solutions to this local model's estimation problem if we express our PDF in the Fourier domain [16].

Currently the orientation correction step is the bottleneck in our algorithm. If we are only interested in the magnitude of the real channel, and thus are indifferent to the orientation of the estimate, our algorithm is very fast — on the order of seconds per slice — and impressively robust across a wide variety of test images. However, our iterative orientation correction step can run for many iterations, taking several minutes per slice, if we require that the orientation be recovered as well. This is also the step least based in the statistical theory we have used to justify the other components of the algorithm. One interesting approach to this problem was suggested by Chang and Xiang [19]: work on the squared complex image and then divide the resulting estimate of ϕ by two at every voxel. This approach is feasible when a polynomial is used to express the value of ϕ, since the division

by two occurs before the estimate is wrapped onto $(-\pi, \pi)$. However, since our algorithm has only an angular estimate of ϕ, division by two is not meaningful and so their idea cannot be transported to our algorithm directly. Regardless, providing a better solution to this aspect of our estimation algorithm would be a useful improvement to our current proposal.

Bibliography

[1] Craig K. Abbey and François O. Bouchud. *Handbook of Medical Imaging*, volume 1, chapter 11 - Modelling Visual Detection Tasks in Correlated Image Noise with Linear Model Observers, pages 629–654. SPIE - The International Society for Optical Engineering, 2000.

[2] C. B. Ahn and Z. H. Cho. A new phase correction method in NMR imaging based on autocorrelation and histogram analysis. *IEEE Trans. Med. Imag.*, 6(1):32–36, March 1987.

[3] M. E. Alexander, R. Baumgartner, A. R. Summers, C. Windischberger, M. Klarhoefer, E. Moser, and R. L. Somorjai. A wavelet-based method for improving signal-to-noise ratio and contrast in MR images. *Magn. Reson. Imaging*, 18:169–180, 2000.

[4] Gilles Aubert and Pierre Kornprobst. *Mathematical problems in image processing : partial differential equations and the calculus of variations.* Springer, 2nd edition, 2006.

[5] Suyash P. Awate and Ross T. Whitaker. Feature-preserving MRI denosing: A nonparametric empirical Bayes approach. *IEEE Trans. Med. Imag.*, 26(9):1242–1255, September 2007.

[6] Paul Bao and Lei Zhang. Noise reduction for magnetic resonance images via adaptive multiscale products thresholding. *IEEE Trans. Med. Imag.*, 22(9):1089–1099, September 2003.

[7] O. E. Barndorff-Nielsen. Adjusted versions of profile likelihood and directed likelihood, and extended likelihood. *Journal of the Royal Statistical Society B*, 56(1):125–140, 1994.

[8] O. E. Barndorff-Nielsen and D. R. Cox. *Inference and Asymptotics.* Monographs on Statistics and Probability. Chapman and Hall, 1994.

[9] Harrison H. Barrett, Craig K. Abbey, and Eric Clarkson. Objective assessment of image quality. III. ROC metrics, ideal observers, and likelihood-generating function. *J. Opt. Soc. Am. A*, 15(6):1520–1535, June 1998.

[10] M. S. Bartlett. Approximate confidence intervals: III. a bias correction. *Biometrika*, 42(1/2):201–204, June 1955.

[11] Debabrata Basu. On the elimination of nuisance parameters. *Journal of the American Statistical Association*, 72(358):355–366, June 1977.

[12] C. Beccari. Typesetting mathematics for science and technology according to ISO 31/XI. *TUGBoat*, 18(1):39–48, 1997.

[13] M. A. Bernstein, D. M. Thomasson, and W.H. Perman. Improved detectability in low signal-to-noise ratio magnetic resonance images by means of a phase-corrected real reconstruction. *Med. Phys.*, 16:813–817, September 1989.

[14] Matthew A. Bernstein and William H. Perman. Least-squares algorithm for phasing MR images. In *Book of Abstracts vol. 2, Sixth Annual Meeting of SMRM*, page 801. Society of Magnetic Resonance in Medicine, August 1987.

[15] J. M. Bonny, M. Zanca, J. Y. Boire, and A. Veyre. T_2 maximum likelihood estimation from multiple spin-echo magnitude images. *Magn. Reson. Med.*, 36:287–293, 1996.

[16] Ronald N. Bracewell. *The Fourier Transform and its Applications.* McGraw Hill, 3rd edition, 2000.

[17] A. E. Burgess. Statistically defined backgrounds: performance of a modified non-prewhitening observer model. *J. Opt. Soc. Am. A*, 11(4):1237–1242, April 1994.

[18] Thomas Cecil, Stanley Osher, and Luminita Vese. Numerical methods for minimization problems constrained to S^1 and S^2. *Journal of Computational Physics*, 198:567–579, 2004.

[19] Zheng Chang and Qing-San Xiang. Nonlinear phase correction with an extended statistical algorithm. *IEEE Trans. Med. Imag.*, 24(6):791–798, June 2005.

[20] Sofia Chavez, Qing-San Xiang, and Li An. Understanding phase maps in MRI: A new cutline phase unwrapping method. *IEEE Trans. Med. Imag.*, 21(8), August 2002.

[21] Scott Daly. *Digital Images and Human Vision*, chapter 14 - The Visible Differences Predictor: An algorithm for the assessment of image fidelity. MIT Press, 1993.

[22] David L. Donoho. De-noising by soft-thresholding. *IEEE Trans. Med. Imag.*, 41(3):613–627, May 1995.

[23] Miguel P. Eckstein, Craig K. Abbey, and François O. Bochud. *Handbook of Medical Imaging*, volume 1, chapter 10 - A Practical Guide to Model Observers for Visual Detection in Synthetic and Natural Noisy Images, pages 593–628. SPIE - The International Society for Optical Engineering, 2000.

[24] Miguel P. Eckstein, Craig K. Abbey, François O. Bochud, Jay L. Bartoff, and James S. Whiting. The effect of image compression in model and human performance. In *Proc. SPIE*, volume 3663, pages 243–252, 1999.

[25] Miguel P. Eckstein, Craig K. Abbey, and James S. Whiting. Human vs model observers in anatomic backgrounds. In *Medical Imaging 1998*, volume 3340 of *Proceedings of the SPIE*, pages 16–26, 1998.

[26] Miguel P. Eckstein, Yani Zhang, and Binh T. Pham. Metrics of medical image quality: task-based model observers vs. image discimination/perceptual difference models. In *Proc. SPIE: Medical Imaging: Image Perception, Observer Performance, and Technology Assessment,*, pages 42–52, 2004.

[27] W. A. Edelstein, G. H. Glover, C. J. Hardy, and R. W. Redington. The intrinsic signal-to-noise ratio in NMR imaging. *Magn. Reson. Med.*, 3:604–618, 1986.

[28] Guido Gehrig, Olaf Kübler, Ron Kikinis, and Ferenc A. Jolesz. Nonlinear anisotropic filtering of MRI data. *IEEE Trans. Med. Imag.*, 11(2):221–232, June 1992.

[29] Bernd Girod. *Digital Images and Human Vision*, chapter 15 - What's wrong with mean-squared error? MIT Press, 1993.

[30] Halfdan Grage and Mikael Akke. A statistical analysis of NMR spectrometer noise. *Journal of Magnetic Resonance*, 162:176–188, 2003.

[31] Mark A. Griswold, Peter M. Jakob, Robin M. Hidemann, mathias Nittka, Valdimir Jellus, Jianmin Wang, Berthold Kiefer, and Axel Hasse. Generalized autocalibrating partially parallel acquisitions (grappa). *Magn. Reson. Med.*, 47:1202–1210, 2002.

[32] Hakon Gudbjartsson and Samuel Patz. The Rician distribution of noisy MRI data. *Magn. Reson. Med.*, 34:910–914, 1995.

[33] R. M. Henkelman and M. J. Bronskill. Artifacts in magnetic resonance imaging. *Reviews of Magnetic Resonance in Medicine*, 2(1):1–126, 1987.

[34] R. Mark Henkelman. Measurement of signal intensities in the presence of noise in MR images. *Med. Phys.*, 12(2):232–233, March 1985.

[35] D. I. Hoult and R. E. Richards. The signal-to-noise ratio of the nuclear magnetic resonance experiment. *Journal of Magnetic Resonance*, 1976.

[36] John D. Kalbfleisch and D. A. Sprott. Application of likelihood methods to models involving large numbers of parameters. *Journal of the Royal Statistical Society B*, 32(2):175–208, 1970.

[37] Kevin F. King and Paul R. Moran. A unified description of NMR imaging, data-collection strategies, and reconstruction. *Medical Physics*, 11(1):1–14, January 1984.

[38] Cheng Guan Koay and Peter J. Basser. Analytically exact correction scheme for signal extraction from noisy magnitude MR signals. *J. Magn. Reson.*, 179:317–322, 2006.

[39] Harold L. Kundel. *Handbook of Medical Imaging*, chapter 18 - Visual Search in Medical Images, pages 837–858. SPIE - The International Society for Optical Engineering, 2000.

[40] V. Kuperman. *Magnetic Resonance Imaging - Physical Principles and Applications.* Academic Press, 2000.

[41] E. L. Lehmann and George Casella. *Theory of Point Estimation.* Spinger, 2nd edition, 1998.

[42] Bruce Levin and Fanhui Kong. Bartlett's bias correction to the profile score function is a saddlepoint correction. *Biometrika*, 77(1):219–221, 1990.

[43] Zhi-Pei Liang. A model-based method for phase unwrapping. *IEEE Trans. Med. Imag.*, 15(6):893–897, December 1996.

[44] Zhi-Pei Liang and Paul C. Lauterbur. *Principles of Magnetic Resonance Imaging.* IEEE Press, 2000.

[45] Nikos K. Logothetis, Jon Pauls, Mark Augath, Torsten Trinath, and Axel Oeltermann. Neurophysiological investigation of the basis of the fMRI signal. *Nature*, 412:150–157, 2001.

[46] James R. MacFall, Norber J. Pelc, and Robert M. Vavrek. Correction of spatially dependent phase shifts for partial Fourier imaging. *Magnetic Resonance Imaging*, 6(2):143–155, 1988.

[47] Luís Miguel Machado and F. Silva Leite. Fitting smooth paths on Riemannian manifolds. *International Journal of Applied Mathematics and Statistics*, 4(J06):25–53, June 2006.

[48] Albert Macovski. Noise in MRI. *Magn. Reson. Med.*, 36:494–497, 1996.

[49] G. McGibney. Phase sensitive reconstruction of MR images. Master's thesis, University of Calgary, 1991.

[50] G. McGibney and M. R. Smith. An unbiased signal-to-noise ratio measure for magnetic resonance images. *Med. Phys.*, 20(4):1077–1078, July 1993.

[51] G. McGibney, M. R. Smith, S. T. Nichols, and A. Crawley. Quantitative evaluation of several partial fourier reconstruction algorithms used in MRI. *Magn. Reson. Med.*, 30:51–59, 1993.

[52] Facundo Mémoli, Guillermo Sapiro, and Stanley Osher. Solving variational problems and partial differentatial equations mapping into general target manifolds. *Journal of Computational Physics*, 195(1):263–292, 2004.

[53] Charles E. Metz. *Handbook of Medical Imaging*, volume 1, chapter 15 - Fundamental ROC Analysis, pages 751–769. SPIE - The International Society for Optical Engineering, 2000.

[54] Jun Miao, Donglai Huo, and David Wilson. Perceptual difference model (case-PDM) for evaluation of MR images: validation and calibration. In *Image perception, observer performance, and technology assessment*, volume 6515 of *Proceedings of the SPIE Medical Imaging 2007:*, page 15. SPIE, SPIE - The International Society for Optical Engineering, 2007.

[55] A. J. Miller and P. M. Joseph. The use of power images to perform quantitative analysis on low SNR MR images. *Magn. Reson. Imaging*, 11:1051–1056, 1993.

[56] Kyle J. Myers. *Handbook of Medical Imaging*, volume 1, chapter 9 - Ideal Observer models of Visual Signal Detection, pages 559–592. SPIE - The International Society for Optical Engineering, 2000.

[57] Kyle J. Myers and Harrison H. Barrett. Addition of a channel mechanism to the ideal-observer model. *J. Opt. Soc. Am. A*, 4(12):2447–2457, December 1987.

[58] D. Noll, D. Nishimura, and A. Macovski. Homodyne detection in magnetic resonance imaging. *IEEE Trans. Med. Imag.*, 10:154–163, 1991.

[59] R. D. Nowak. Wavelet-based Rician noise removal for magnetic resonance imaging. *IEEE Trans. Image Processing*, 8(10):1408–1418, October 1999.

[60] Stanley J. Osher. *Level Set and Dynamic Implicit Surfaces.* Spinger, 2002.

[61] Subok Park. Efficiency of the human observer detecting random signals in random backgrounds. *J. Opt. Soc. Am. A*, 22(1):3–16, January 2005.

[62] P. Perona and J. Malik. Scale-space and edge detection using anisotropic diffusion. *Pattern Analysis and Machine Intelligence, IEEE Transactions on , Vol.12, Iss.7, 1990*, 12(7):629–639, 1990.

[63] Pietro Perona. Orientation diffusions. *IEEE Transactions on Image Processing*, 7(3):457–467, March 1998.

[64] Aleksandra Pižuica, Wilfried Philips, Ignace Lemahieu, and Marc Acheroy. A versatile wavelet domain noise filtration technique for medical imaging. *IEEE Trans. Med. Imag.*, 22(3):323–331, March 2003.

[65] Klass P. Pruessmann, Markus Weiger, Markus B. Scheidegger, and Peter Boesiger. SENSE: Sensitivity encoding for fast MRI. *Magn. Reson. Med.*, 42:952–962, 1999.

[66] S. O. Rice. Mathematical analysis of random noise. In *Selected Papers on Stochastic Processes*, volume 23, chapter 4, pages 133–294. Dover, 1954.

[67] J. P. Rolland and H. H. Barrett. Effect of random background inhomogeneity on observer detection performance. *J. Opt. Soc. Am. A*, 9(5):649–658, May 1992.

[68] Kyle A. Salem, Jonathan S. Lewin, Andrik J. Aschoff, Jeffrey L. Duerk, and David L. Wilson. Validation of a human vision model for image quality evaluation of fast interventional magnetic resonance imaging. *Journal of Electronic Imaging*, 11(2):224–235, April 2002.

[69] John F. Schenck. The role of magnetic susceptibility in magnetic resonance imaging: MRI magnetic compatibility of the first and second kinds. *Medical Physics*, 23(6):815–850, July 1996.

[70] J. Sijbers, A. J. den Dekker, E. Raman, and D. Van Dyck. Parameter estimation for magnitude MR images. *International Journal of Imaging Systems and Technology*, 10(2):109–114, 1999.

[71] Jan Sijbers and Arnold J. den Dekker. Maximum likelihood estimation of signal amplitude and noise variance from MR data. *Magn. Reson. Med.*, 51:586–594, 2004.

[72] Jan Sijbers, Arnold J. den Dekker, Paul Scheunders, and Dirk Van Dyck. Maximum-likelihood estimation of Rician distribution parameters. *IEEE Trans. Med. Imag.*, 17(3):357–361, June 1998.

[73] J. F. L. Simmons and B. G. Stewart. Point and inverval estimation of the true unbiased degree of linear polarization in the present of low signal-to-noise ratios. *Astronomy and Astrophysics*, 142:100–106, 1985.

[74] D. K. Sodickson and W. J. Manning. Simultaneous acquisition of spatial harmonics (SMASH): Fast imaging with radiogrequency coil arrays. *Magn. Reson. Med.*, 38:591–603, 1997.

[75] B. G. Stewart and J. F. L. Simmons. Rician envelope estimation and confidence intervals in low signal/noise levels. *Electronics Letters*, 23(16):832–834, July 1987.

[76] Edward Susko, Michael J. Bronskill, Simon J. Graham, and Robert J. Tibshirani. Estimation of relaxation time distributions in magnetic resonance imaging. *The Canadian Journal of Statistics*, 29(3):379–394, 2001.

[77] Bei Tang, Guillermo Sapiro, and Vincent Caselles. Color image enhancement via chromaticity diffusion. *IEEE Transactions on Image Processing*, 10(5):701–707, May 2001.

[78] Stefan Thesen, Oliver Heid, Edgar Mueller, and Lothar R. Schad. Prospective acquisition correction for head motion with image-based tracking for real-time fMRI. *Magn. Reson. Med.*, 44:457–465, 2000.

[79] M. Dylan Tisdall and M. Stella Atkins. MRI denoising via phase error estimation. In *Medical Imaging 2005*, volume 5747 of *Proceedings of the SPIE*, pages 646–654, 2005.

[80] M. Dylan Tisdall and M. Stella Atkins. Using human and model performance to compare MRI reconstructions. *IEEE Trans. Med. Imag.*, 25(11):1510–1517, November 2006.

[81] M. Dylan Tisdall and M. Stella Atkins. Perception of dim targets on dark backgrounds in MRI. In *Medical Imaging 2007*, volume 6515 of *Proceedings of the SPIE*, page 651513, 2007.

[82] M. Dylan Tisdall, M. Stella Atkins, and Richard A. Lockhart. Maximum likelihood estimators in magnetic resonance imaging. In *Information Processing in Medical Imaging*, volume 4584 of *Lecture Notes in Computer Science*, pages 434–445. Springer, 2007.

[83] André J. W. van der Kouwe, Thomas Benner, and Anders M. Dale. Real-time rigid body motion correction and shimming using cloverleaf navigators. *Magn. Reson. Med.*, 56:1019–1032, 2006.

[84] Luminita A. Vese and Stanley J. Osher. Numerical methods for p-harmonic flows and applications to image processing. *SIAM Journal on Numerical Analysis*, 40(6):2085–2104, 2002.

[85] M. T. Vlaardingerbroek and J. A. den Boer. *Magnetic Resonance Imaging.* Springer-Verlag, 2nd edition, 1999.

[86] Grace Wahba. *Spline models for observational data.* Society for Industrial and Applied Mathematics, 1990.

[87] Andrew B. Watson and Joshua A. Solomon. Model of visual contrast gain control and pattern masking. *J. Opt. Soc. Am. A*, 14(9):2379–2391, September 1997.

[88] J. Weaver, Y. Xu, D. Healy, and J. Driscoll. Filtering MR images in the wavelet transform domain. *Magn. Reson. Med.*, 21:288–295, 1991.

[89] Kenneth P. Whittal and Alexander L. MacKay. Quantitative interpretation of NMR relaxation data. *Journal of Magnetic Resonance*, 84:134–152, 1989.

[90] Wolfram Research Inc. *Mathematica Version 6.0.* Wolfram Research Inc., 2007.

[91] J. C. Wood and K. Johnson. Wavelet-denoising of complex magnetic resonance images. In *Bioengineering Conference. Proceedings of the IEEE 24th Annual Northeast*, pages 32–34, 1998.

[92] G. Wyszecki and W. S. Stiles. *Color Science: Concepts and Methods, Quantitative Data and Formulae.* John Wiley and Sons, Inc., second edition, 1982.

[93] E. A. Zerhouni, D. M. Parish, W. J. Rogers, A. Yang, and E. P. Shapiro. Human heart: tagging with MR imaging – a method for noninvasive assessment of myocardial motion. *Radiology*, 169:59–63, 1988.

[94] Yani Zhang, Binh T. Pham, and Miguel P. Eckstein. The effect of non-linear human visual system components on linear model observers. In *Proc. SPIE: Medical Imaging: Image Perception, Observer Performance, and Technology Assessment,*, pages 31–41. SPIE - The International Society for Optical Engineering, 2004.

[95] Reza Aghaeizadeh Zoroofi, Yoshinobu Sato, Shinichi Tamura, and Hiroaki Naito. MRI artifact cancellation due to rigid motion in the imaging plane. *IEEE Trans. Med. Imag.*, 15(6):768–784, December 1996.